UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# MOLECULAR CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF ORF P1192R FROM AFRICAN SWINE FEVER VIRUS

João Nuno Santos Coelho

ORIENTADOR: Doutor José Alexandre da Costa Perdigão e Cameira Leitão

CO-ORIENTADOR: Prof. Doutor Fernando António da Costa Ferreira

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Veterinárias, na especialidade de Ciências Biológicas e Biomédicas.

2016

UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# MOLECULAR CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF ORF P1192R FROM AFRICAN SWINE FEVER VIRUS

João Nuno Santos Coelho

ORIENTADOR: Doutor José Alexandre da Costa Perdigão e Cameira Leitão

CO-ORIENTADOR: Prof. Doutor Fernando António da Costa Ferreira

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Veterinárias, na especialidade de Ciências Biológicas e Biomédicas.

JÚRI:
Presidente: Prof. Doutor Rui Manuel de Vasconcelos e Horta Caldeira
Vogais:
Doutora Linda Kathleen Dixon
Prof. Doutor Carlos Manuel Lopes Vieira Martins
Prof. Doutora Aida Maria da Conceição Esteves Simões
Prof. Doutora Ana Isabel Simões Pereira Duarte
Prof. Doutor Fernando António da Costa Ferreira
Doutor José Alexandre da Costa Perdigão e Cameira Leitão

2016

# ACKNOWLEDGMENTS

There are many people to whom I'm thankful and which have contributed, in one way or another, to the progression and conclusion of this work.

The most important contributor was, without a doubt, my supervisor Alexandre Leitão. I'm extremely grateful to him for allowing me to follow my own ideas, while challenging me with his own views on the matters at hand, and for giving me the freedom I needed to achieve the goals that were initially set. I'm grateful not only for the helpful discussions, for his support and for all the knowledge he shared, but also for the things he didn't know and that I had to learn and discover by myself. Thanks to him, I believe I was able to truly learn how to conduct scientific research independently, which is one of the objectives of a true PhD.

I also acknowledge my co-supervisor Fernando Ferreira for all his contributions to the progression of my doctoral studies and to the elaboration of this document, and to Professor Carlos Martins for his advice and for allowing me to conduct my PhD studies in the laboratory under his tutelage.

To my colleagues during my time at the Infectious Diseases Laboratory, Raquel Portugal, Margarida Simões, Ferdinando Freitas and Gonçalo Frouco, I'm grateful for their help, collaboration and for useful work discussions. I also acknowledge Maria de Jesus Silva, Rui Vieira, Conceição Trigo and Paula Viana for all the important backstage work.

To my friends (and in alphabetic order of the surname, in order to avoid overgrown egos) Silvia Almeida, Afonso Basto, Clara Cartaxeiro, Andreia Ferreira, Samuel Francisco, Carla Mottola, Sofia Nolasco, Dulce Santos, Alexandra Tavares and Sara Zúquete, I'm thankful for their infinite patience and their strange and endless capacity to cope with my annoying presence. I was able to learn something from each of them and their friendship was very important during this endeavour.

For technical help, advice and/or biological material, I'm grateful to Marina Badenes, Caroline Austin, Luísa Cyrne, Rodrigo Cunha, Ian Cheeseman, Ana Cecília Seixas, Günther Keil, Michael Parkhouse, Sílvia Correia, Ângela Lopes and Aida Esteves. I must also thank my former supervisor, Lisete Fernandes, since it is likely that without all that I learned from her I would have never been able to accomplish the main objective of this PhD.

Finally, a last word of acknowledgment goes to Cátia, Diana and Sofia, my wife and daughters, for understanding my absence even when I was present and for being patient and helping me to see this through.

# FUNDING

The present work was funded by PhD fellowship SFRH/BD/48654/2008 and grant PTDC/CVT/105630/2008 – ASFTOPO from Fundação para a Ciência e Tecnologia, and by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 311931, ASFORCE.


# ABSTRACT

African swine fever virus (ASFV) is a nucleo-cytoplasmic large DNA arbovirus and the single member of the family Asfarviridae. It infects soft ticks of the genus *Ornithodoros* as well as all members of the family Suidae, representing a global threat for pig husbandry for which there is currently no effective vaccine or treatment. Since the ASFV viral cycle is mainly cytoplasmic, it has been found/predicted to code for many components of the replicative and transcriptional machineries. Of these, and based in sequence homologies, a putative type II DNA topoisomerase-coding ORF (P1192R) was identified in the ASFV genome. DNA topoisomerases are enzymes that modulate the topological state of DNA molecules. They are ubiquitous and essential, participating in processes such as DNA replication, recombination and repair and also in transcription. Since ASFV has a large linear genome, with 170 to 190 kbp depending on the isolate, containing terminal inverted repeats and covalently closed ends, a type II topoisomerase may be indispensable for viral replication and/or transcriptional events. The main objectives of this work were to deepen the study on ORF P1192R and determine if it indeed codes for a type II DNA topoisomerase and, if so, to characterize its activity. Bioinformatics and phylogenetic analyses showed that ORF P1192R is highly conserved among the fourteen ASFV isolates analyzed and, although its amino acid sequence clearly diverges from other type II topoisomerases, the structural organization is preserved and conserved motifs and domains essential for activity are present. Transient expression of GFP-pP1192R in COS-7 cells revealed an exclusively cytoplasmic distribution of the protein, which remained unaltered by treatment with leptomycin B. Using Vero cells or swine macrophages infected with ASFV isolate Ba71V or L60, respectively, expression of pP1192R was observed in the late phase of infection, co-localizing with the viral factories, where the bulk of viral replication and transcription occurs. Heterologous expression of pP1192R in *Saccharomyces cerevisiae* demonstrated that it functionally complements a *top2* thermo-sensitive mutation and that it exhibits ATP-dependent decatenation activity. The purified recombinant pP1192R was found to efficiently decatenate kDNA and to processively relax supercoiled plasmid DNA, which are characteristics of a type II topoisomerase. The optimal requirements in terms of pH, temperature and salt, divalent ions and ATP concentrations for pP1192R activity *in vitro* were determined and its sensitivity to a panel of topoisomerase poisons and inhibitors was tested. Our results indicate that P1192R may be a target for studying, and possibly controlling, ASFV transcription and replication.

Keywords: African swine fever virus, ASFV, DNA topoisomerase, P1192R

# RESUMO

O vírus da peste suína africana (VPSA) é um arbovírus icosaédrico núcleo-citoplasmático de DNA de cadeia dupla, classificado no género *Asfivirus* da família Asfarviridae, da qual é o único membro conhecido. Este vírus infecta carraças do género *Ornithodoros* assim como todos os membros da família Suidae, constituindo uma ameaça global para a suinicultura para a qual não existe actualmente qualquer vacina ou tratamento. A prevenção da peste suína africana é feita através de medidas que visam reduzir o risco de introdução de animais ou produtos de origem animal infectados em regiões livres da doença, enquanto o controlo de um surto se baseia exclusivamente em medidas que incluem o abate sanitário de todos os animais susceptíveis na área do foco e a proibição de movimentos e comercialização de animais.

Embora o VPSA tenha sido inicialmente descrito como um vírus com replicação exclusivamente citoplasmática, actualmente sabe-se que o núcleo da célula hospedeira é indispensável na fase inicial da infecção. Contudo, a grande maioria do ciclo infeccioso ocorre no citoplasma da célula infectada, não sendo por isso surpreendente que, das 150 a 167 grelhas de leitura aberta (ORF, do inglês “open reading frame”) identificadas no genoma do VPSA, algumas codifiquem para componentes das maquinarias de replicação e de transcrição. Dentre estas, prevê-se, com base em homologia de sequências aminoacídicas, que a ORF P1192R codifique para uma topoisomerase de DNA do tipo II.

As topoisomerases de DNA estão presentes em todas as células e são responsáveis pela modulação do estado topológico do DNA, estado esse que se altera durante processos como a replicação, a recombinação e a reparação do DNA, assim como a transcrição, e dos quais resultam torções das moléculas de DNA que, não sendo resolvidas, podem comprometer a integridade genómica e consequentemente a viabilidade celular. Todas as topoisomerases exercem a sua actividade através da criação de quebras no DNA devido ao ataque nucleofílico de um resíduo de tirosina catalítico ao esqueleto fosfodiéster da molécula de DNA, gerando-se assim uma ligação fosfotirosina covalente. As topoisomerases são classificadas em dois tipos, tendo por base a forma como quebram a molécula de DNA: as topoisomerases do tipo I, cuja actividade é independente de ATP e que geram quebras em cadeia única no DNA, facilitando assim o desenrolamento; e as topoisomerases do tipo II, que necessitam de ATP para gerar uma quebra nas duas cadeias do DNA, através da qual fazem passar uma dupla cadeia intacta. Considerando que o VPSA tem um genoma linear de grandes dimensões, com 170 a 190 quilopares de bases dependendo do isolado, e que contém repetições terminais

invertidas fechadas covalentemente, uma topoisomerase do tipo II pode efectivamente ser essencial para eventos de replicação e/ou transcrição virais.

Os objectivos centrais deste trabalho foram os seguintes: (i) realização de um estudo bioinformático e filogenético aprofundado da ORF P1192R do VPSA; (ii) estudo da proteína codificada por esta ORF (pP1192R), através da sua clonagem, expressão em sistema heterólogo, purificação da proteína recombinante e caracterização *in vitro* da sua actividade; (iii) determinação do efeito sobre a actividade da proteína recombinante dum painel de compostos químicos descritos como sendo inibidores de topoisomerases; (iv) identificação dos níveis de expressão e da localização intracelular da pP1192R em células infectadas pelo VPSA, a diferentes tempos de infecção; (v) avaliação do efeito de mutações dirigidas em resíduos ou motivos identificados como reguladores da actividade enzimática ou localização subcelular da pP1192R, tendo por base a informação gerada nos estudos bioinformáticos acima mencionados.

A ORF P1192R do isolado L60 do VPSA foi amplificada por PCR e clonada e a sua sequência nucleotídica foi determinada e utilizada em análises bioinformáticas e filogenéticas. Verificou-se que esta ORF é altamente conservada entre os catorze isolados do VPSA cujo genoma se encontrava disponível nas bases de dados e, embora a sua sequência aminoacídica seja claramente divergente das de outras topoisomerases do tipo II incluídas neste estudo, quer sejam elas de origem procariota, eucariota ou viral, a organização estrutural da proteína está preservada e estão presentes motivos e domínios conservados que são essenciais para a actividade enzimática. O estudo da localização celular da pP1192R iniciou-se com a construção de plasmídeos quiméricos para a expressão da pP1192R em fusão com a proteína verde fluorescente (GFP) ou com uma variante vermelha (RFP). Transfectaram-se transientemente células de linha COS-7 com estas construções tendo-se observado que a proteína de fusão se distribuía exclusivamente pelo citoplasma. Esta distribuição não foi alterada após tratamento com leptomicina B que bloqueia uma das vias de exportação de proteínas do núcleo. Já a infecção das células a expressarem GFP-pP1192R com um isolado do VPSA adaptado a células Vero (Ba71V) induziu uma redistribuição da proteína de fusão, deixando de estar homogeneamente distribuída pelo citoplasma para estar principalmente concentrada nas fábricas virais a partir das 8 horas pós-infecção. Utilizando células de linha Vero infectadas com o isolado Ba71V, utilizado como modelo de infecção, ou macrófagos derivados de monócitos de sangue periférico de suíno (células alvo do vírus na infecção natural) infectados com o isolado virulento L60, e utilizando um soro anti-pP1192R produzido no decurso destes trabalhos, foi possível constatar que a pP1192R viral é produzida

na fase intermédia/tardia da infecção (observável a partir das 6/8 horas pós-infecção) e que acumula nas fábricas virais ao longo da infecção.

A expressão em sistema heterólogo da pP1192R iniciou-se num sistema procariota, baseado em *Escherichia coli*, mas embora tenha sido possível obter proteína recombinante em grandes quantidades, a sua purificação só foi conseguida recorrendo a agentes desnaturantes, impedindo a obtenção de proteína activa. Assim, avançou-se para um novo sistema de expressão baseado na levedura *Pichia pastoris* que apresenta diversas vantagens sobre o anterior, nomeadamente o facto de ser um sistema eucariota e por isso mais semelhante ao contexto em que a pP1192R é expressa em condições naturais. Contudo, neste sistema não foi possível obter proteína recombinante e o sistema foi abandonado. Tentou-se por fim a expressão heteróloga na levedura *Saccharomyces cerevisiae*. Neste organismo, a utilização das estirpes JCW26 e SD117 que contêm uma mutação termo-sensível no gene que codifica para a topoisomerase do tipo II endógena, permitiu demonstrar, quer *in vivo* através da complementação da mutação termo-sensível, quer *in vitro* recorrendo a ensaios funcionais de decatenação, que a pP1192R é efectivamente uma topoisomerase do tipo II funcional. Utilizando ainda *S. cerevisiae* como sistema de expressão, foi possível obter e purificar pP1192R recombinante para caracterização da sua actividade em ensaios funcionais *in vitro*. Observou-se que a pP1192R é capaz de relaxar DNA superenrolado, de decatenar DNA catenado e, quando em elevadas concentrações, de catenar DNA plasmídico, não tendo sido detectada actividade de superenrolamento de DNA relaxado. Determinaram-se também as condições óptimas de funcionamento em termos de temperatura, pH e concentrações de sal (NaCl ou KCl), ATP ou iões divalentes ($Mg^{2+}$, $Mn^{2+}$, $Zn^{2+}$, $Cu^{2+}$ e $Ca^{2+}$), que foram posteriormente utilizadas para avaliar a sensibilidade da pP1192R recombinante a um painel de inibidores de topoisomerases, entre os quais se incluem drogas frequentemente utilizadas como agentes antimicrobianos ou antitumorais. Dos compostos testados, aqueles para os quais foram obtidos resultados mais promissores, i.e., os que revelaram níveis de inibição mais elevados, foram a coumermicina A1, a doxorubicina, a amsacrina e a genisteína. Pelo contrário, as quinolonas, normalmente utilizadas como antibióticos visando infecções provocadas por organismos procariotas, foram dos compostos com menor eficácia.

Em suma, os resultados deste trabalho indicam que a ORF P1192R é um alvo promissor para o estudo e, eventualmente, o controlo dos processos replicativos e transcricionais do vírus da peste suína africana.

Palavras-chave: vírus da peste suína africana, VPSA, DNA topoisomerase, P1192R

# TABLE OF CONTENTS

**CHAPTER 1: INTRODUCTION** ..... **1**

**1.1 AFRICAN SWINE FEVER** ..... **3**

***1.1.1 Historical synopsis*** ..... **3**

***1.1.2 Hosts of African swine fever virus*** ..... **4**

***1.1.3 Transmission of African swine fever*** ..... **6**

***1.1.4 Molecular epidemiology*** ..... **6**

***1.1.5 Morphology of ASFV*** ..... **8**

***1.1.6 Replication and transcription of ASFV*** ..... **11**

**1.2 DNA TOPOISOMERASES** ..... **15**

***1.2.1 Brief considerations on the discussion of DNA topoisomerases and topological problems*** ..... **15**

***1.2.2 Overview of DNA topoisomerases*** ..... **15**

***1.2.3 The catalytic reaction of type II DNA topoisomerases*** ..... **19**

***1.2.4 The biological roles of type II DNA topoisomerases*** ..... **23**

***1.2.5 Inhibition of type II DNA topoisomerases*** ..... **26**

**1.3 AIM OF THE WORK** ..... **28**

**1.4 REFERENCES** ..... **29**

**CHAPTER 2: BIOINFORMATICS AND PHYLOGENETIC ANALYSIS OF ORF P1192R OF AFRICAN SWINE FEVER VIRUS** ..... **45**

**2.1 INTRODUCTION** ..... **47**

**2.2 MATERIALS AND METHODS** ..... **49**

***2.2.1 Bacterial strains and transformation*** ..... **49**

***2.2.2 Primers used*** ..... **49**

***2.2.3 DNA amplification by PCR for cloning purposes*** ..... **50**

***2.2.4 Agarose gel electrophoresis*** ..... **51**

***2.2.5 DNA purification from agarose slices*** ..... **51**

***2.2.6 Plasmid DNA purification*** ..... **52**

***2.2.7 DNA quantification*** ..... **52**

***2.2.8 Hydrolysis and ligation of DNA fragments*** ..... **52**

***2.2.9 Cloning and sequencing of P1192R from ASFV/L60 and ASFV/E75*** ..... **52**

***2.2.10 Colony PCR*** ..... **53**

***2.2.11 Bioinformatics analyses*** ..... **53**

***2.2.12 Sequence alignment and phylogenetic analyses*** ..... **54**

**2.3 RESULTS** ..... **58**

***2.3.1 Determination of the sequence from L60's ORF P1192R and its analysis using bioinformatics tools*** ..... **58**

***2.3.2 Phylogenetic analyses of P1192R among ASFV isolates*** ..... **64**

***2.3.3 Phylogenetic analysis of viral type II topoisomerases*** ..... **68**

***2.3.4 Phylogenetic distribution of large DNA viruses sequences*** ..... **69**

***2.3.5 Comprehensive phylogenetic analysis of type II topoisomerases*** ..... **73**

**2.4 DISCUSSION** ..... **75**

**2.5 REFERENCES** ..... **78**

**CHAPTER 3: STUDIES ON THE EXPRESSION OF P1192R IN A CELLULAR CONTEXT 83**

**3.1 INTRODUCTION ... 85**
**3.2 MATERIALS AND METHODS ... 86**
***3.2.1 Mammalian cell lines ... 86***
***3.2.2 Viruses and infections ... 86***
***3.2.3 Cloning of P1192R from ASFV/L60 ... 87***
***3.2.4 Electroporation ... 88***
***3.2.5 Immunofluorescence ... 88***
***3.2.6 Topoisomerase II retention assay ... 89***
***3.2.7 Immunoblotting ... 90***
***3.2.8 Antibodies and antisera ... 90***
**3.3 RESULTS ... 91**
***3.3.1 Subcellular localization of GFP-pP1192R ... 91***
***3.3.2 Immunofluorescence analysis of viral pP1192R localization and expression ... 95***
***3.3.3 Analysis of viral pP1192R by Western blot ... 101***
**3.4 DISCUSSION ... 101**
**3.5 REFERENCES ... 102**

**CHAPTER 4: HETEROLOGOUS EXPRESSION AND *IN VITRO* FUNCTIONAL CHARACTERIZATION OF ASFV ORF P1192R ... 107**

**4.1 INTRODUCTION ... 109**
***4.1.1 Recombinant protein expression in Escherichia coli ... 109***
***4.1.2 Recombinant protein expression in yeast ... 111***
***4.1.3 In vitro assays for evaluation of topoisomerase activity ... 114***
**4.2 MATERIALS AND METHODS ... 116**
***4.2.1 Primers used ... 116***
***4.2.2 Hydrolysis and ligation of DNA fragments ... 117***
***4.2.3 Cloning of P1192R from ASFV/L60 ... 118***
***4.2.4 Bacterial strains and transformation ... 120***
***4.2.5 Bacterial cultures for protein expression ... 120***
***4.2.6 Expression and purification of recombinant pP1192R from Escherichia coli ... 120***
***4.2.7 Maintenance of Pichia pastoris ... 121***
***4.2.8 Transformation and selection of Pichia pastoris ... 121***
***4.2.9 DNA extraction from yeast cells ... 122***
***4.2.10 Protein extraction from yeast cells for immunoblotting ... 123***
***4.2.11 Confirmation of Pichia integrants by PCR ... 123***
***4.2.12 Expression of P1192R in Pichia pastoris ... 123***
***4.2.13 Maintenance of Saccharomyces cerevisiae ... 124***
***4.2.14 Transformation of Saccharomyces cerevisiae ... 124***
***4.2.15 Production of polyclonal anti-pP1192R serum using purified protein expressed in Saccharomyces cerevisiae ... 125***
***4.2.16 Complementation assays ... 125***
***4.2.17 Fluorescence microscopy of Saccharomyces cerevisiae ... 126***
***4.2.18 Preparation of soluble protein extracts from Saccharomyces cerevisiae ... 126***
***4.2.19 Immunoblotting ... 127***
***4.2.20 Expression and purification of functional recombinant pP1192R from Saccharomyces cerevisiae ... 127***

*4.2.21 Assays of type II DNA topoisomerase activity* ..... 128
**4.3 RESULTS** ..... 129
*4.3.1 Heterologous expression of P1192R in Escherichia coli* ..... 129
*4.3.2 Purification of pP1192R expressed in E. coli and in vitro activity assays* ..... 132
*4.3.3 Cloning of P1192R, transformation and selection of Pichia pastoris* ..... 132
*4.3.4 Attempts to express recombinant protein in Pichia pastoris* ..... 134
*4.3.5 Heterologous expression of P1192R in Saccharomyces cerevisiae* ..... 135
*4.3.6 Expression and purification of pP1192R* ..... 136
*4.3.7 Production of polyclonal anti-pP1192R serum using recombinant protein expressed in S. cerevisiae* ..... 138
*4.3.8 Expression and purification of Top2p* ..... 139
*4.3.9 Functional complementation of Top2p thermo-sensitive S. cerevisiae mutants* ..... 140
*4.3.10 In vitro demonstration of pP1192R topoisomerase II activity* ..... 144
*4.3.11 Purification of functional recombinant pP1192R* ..... 145
*4.3.12 Characterization of pP1192R activity in vitro* ..... 147
*4.3.13 Inhibition of pP1192R by using anti-topoII compounds* ..... 153
**4.4 DISCUSSION** ..... 156
**4.5 BIBLIOGRAPHY** ..... 160

**CHAPTER 5: CONCLUDING REMARKS** ..... 167

**5.1 CHARACTERIZATION OF PP1192R** ..... 169
**5.2 THE ROLE OF PP1192R DURING VIRAL INFECTION** ..... 171
**5.3 P1192R AS A POSSIBLE TARGET FOR ANTIVIRAL THERAPY AND VACCINE DESIGN** ..... 173
**5.4 BIBLIOGRAPHY** ..... 175

# INDEX OF FIGURES

**Figure 1.** Structure and protein composition of ASFV particle. ..... 9
**Figure 2.** Accumulation kinetics for immediately early (blue, I215L), early (green, I73R), intermediate (yellow, I226R) and late (red, I329L) transcripts throughout ASFV infection. ..... 12
**Figure 3.** Schematic representation of the self-priming model for replication of the ASFV genome, adapted from the model proposed for poxviruses..... 14
**Figure 4.** Roles of DNA topoisomerases in maintaining cellular DNA topology. ..... 19
**Figure 5.** Structure of type II DNA topoisomerases. ..... 21
**Figure 6.** Mechanism of strand passage by type II DNA topoisomerases. ..... 23
**Figure 7.** Topological problems associated with the progression of the replication machinery..... 24
**Figure 8.** Catalytic inhibitors of type II topoisomerases. ..... 26
**Figure 9.** ORF P1192R from ASFV isolate L60. ..... 59
**Figure 10.** Phobius prediction of transmembrane domains in P1192R amino acid sequence. ..... 61
**Figure 11.** Functional motifs and domains in the amino acid sequence of P1192R. ..... 62
**Figure 12.** Predicted structure of pP1192R, in its dimer form. ..... 63
**Figure 13.** Phylogenetic analysis of ASFV P1192R. ..... 65
**Figure 14.** Maximum-likelihood phylogenetic analysis of type II topoisomerases from large DNA viruses. ..... 68
**Figure 15.** Maximum-likelihood phylogenetic analysis of RNA polymerase from large DNA viruses. ..... 69
**Figure 16.** Maximum-likelihood phylogenetic analysis of DNA polymerase from large DNA viruses. ..... 71
**Figure 17.** Maximum-likelihood phylogenetic analysis of primase from large DNA viruses. ..... 72
**Figure 18.** Maximum-likelihood phylogenetic analysis of PCNA homologs from large DNA viruses. ..... 73
**Figure 19.** Maximum-likelihood phylogenetic analysis of type II topoisomerases from several domains of the tree of life. ..... 74
**Figure 20.** Pairwise alignment of pP1192R and Top2p. ..... 76
**Figure 21.** Evaluation of the cellular distribution of GFP-pP1192R. ..... 92
**Figure 22.** Effect of leptomycin B on the cellular distribution of GFP-pP1192R. ..... 93
**Figure 23.** Cellular distribution of GFP-pP1192R upon viral infection. ..... 94
**Figure 24.** Recognition of recombinant pP1192R by the anti-pP1192R serum. ..... 95
**Figure 25.** Cellular distribution of viral pP1192R during Ba71V infection in Vero cells. ..... 96
**Figure 26.** Cellular distribution of viral pP1192R during L60 infection in pig macrophages. ..... 97
**Figure 27.** Expression of P1192R in the presence of AraC. ..... 98
**Figure 28.** Expression of vp32 in the presence of AraC. ..... 99
**Figure 29.** Expression of vp72 in the presence of AraC. ..... 100
**Figure 30.** Detection of viral pP1192R by Western blot. ..... 101
**Figure 31.** Regulation of protein expression in the pET system. ..... 111
**Figure 32.** Topoisomerase activity assays. ..... 115
**Figure 33.** Solubilization of recombinant pP1192R through the use of detergents. ..... 131
**Figure 34.** Solubilization of recombinant pP1192R through the use of fos-choline-12. ..... 132
**Figure 35.** Time-course of pP1192R expression in *Saccharomyces cerevisiae*. ..... 135
**Figure 36.** Preliminary purification of recombinant pP1192R_6xHis expressed in *S. cerevisiae*. ..... 136
**Figure 37.** *In vitro* decatenation assay using purified pP1192R. ..... 137
**Figure 38.** Western blot testing of the anti-pP1192R polyclonal serum. ..... 139
**Figure 39.** *In vitro* decatenation assay using purified Top2p. ..... 140

**Figure 40.** Schematic representation of the recombinant proteins expressed in complementation assays.....141
**Figure 41.** Localization of GFP-fusion constructs in *S. cerevisiae* cells.....142
**Figure 42.** Functional complementation of a yeast *top2* temperature-sensitive mutation by P1192R. .....143
**Figure 43.** SDS-PAGE analysis of total protein yeast extracts. .....144
**Figure 44.** Decatenation activity *in vitro* assay using total protein yeast extracts.....145
**Figure 45.** Purification of functional recombinant pP1192R. .....146
**Figure 46.** Determination of pP1192R's concentration to be used in *in vitro* assays for the characterization of its activity. .....147
**Figure 47.** The effect of NaCl (A) or KCl (B) on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R. .....148
**Figure 48.** The effect of pH on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R. 149
**Figure 49.** The effect of temperature on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R. .....149
**Figure 50.** ATP dependence and the effect of ATP concentration on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.....150
**Figure 51.** The effect of divalent ions on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R. .....151
**Figure 52.** Time course of the relaxation of supercoiled DNA by pP1192R. .....152
**Figure 53.** Testing the supercoiling ability of pP1192R.....153
**Figure 54.** Inhibition of decatenation activity of pP1192R. .....154
**Figure 55.** Testing the effect of topotecan on pP1192R activity. .....154
**Figure 56.** Inhibition of decatenation activity of pP1192R by quinolones.....155

# INDEX OF TABLES

**Table 1.** Primers used in this chapter. .....50
**Table 2.** Topoisomerase sequences used for phylogenetic analyses .....55
**Table 3.** GC content of P1192R and other ORFs. .....60
**Table 4.** DNA and amino acid identity percentages among P1192R from fourteen ASFV isolates. ....67
**Table 5.** Components of replicative or transcriptional machineries present in the genome of large DNA viruses. .....70
**Table 6.** Primers used in this chapter. .....88
**Table 7.** Primers used in this chapter. .....116
**Table 8.** Lowest effective doses for topoisomerase inhibitors tested in this work. .....155
**Table 9.** Comparison of optimal conditions for purified eukaryotic type II topoisomerases. .....159

# INDEX OF ABBREVIATIONS

| | |
|---|---|
| aa | amino acid(s) |
| ADP | adenosine 5′-diphosphate |
| ADPNP | 5'-adenylyl-beta,gamma-imidodiphosphate |
| AraC | cytosine arabinoside |
| ASF | African swine fever |
| ASFV | African swine fever virus |
| ATCC | American type culture collection |
| ATCV-1 | *Acanthocystis turfacea Chlorella* virus 1 |
| ATP | adenosine 5′-triphosphate |
| bp | base pair(s) |
| BSA | bovine serum albumin |
| CAA | casamino acids |
| CCR | conserved central region |
| CD2 | T-cell surface antigen CD2 |
| CRM-1 | chromosome region maintenance 1 protein homolog |
| CSF | classical swine fever |
| CTD | C-terminal domain |
| DAPI | 4′,6-diamidino-2-phenylindole |
| dATP | 2'-deoxyadenosine 5'-triphosphate |
| dCTP | 2'-deoxycytidine 5'-triphosphate |
| dGTP | 2'-deoxyguanosine 5'-triphosphate |
| DISC | disabled infectious single cycle |
| DMEM | Dulbecco’s modified Eagle medium |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| DNApol | DNA polymerase |
| dNTP | deoxynucleotide |
| dTTP | 2'-deoxythymidine 5'-triphosphate |
| ECACC | European collection of cell cultures |
| EDTA | ethylenediaminetetraacetic acid |
| EGTA | ethylene glycol-bis(2-aminoethylether)-*N,N,N′,N′*-tetraacetic acid |
| ER | endoplasmic reticulum |
| FBS | fetal bovine serum |
| GFP | green fluorescent protein |
| GHKL | gyrase, Hsp90, histidine kinase, MutL |
| GTP | guanosine 5'-triphosphate |
| HcDNAV | *Heterocapsa circularisquama* DNA virus |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| hpi | hour(s) post-infection |
| HRP | horseradish peroxidase |
| IIV-6 | invertebrate iridescent virus 6 |

| | |
|---|---|
| kbp | kilobase pair(s) |
| LB | lysogeny broth |
| LMB | leptomycin B |
| LVR | left variable region |
| ME | minimum evolution |
| MGF | multigene family |
| ML | maximum likelihood |
| MOI | multiplicity(ies) of infection |
| MP | maximum parsimony |
| MSA | multiple sequence alignment |
| NCLDV | nucleo-cytoplasmic large DNA virus(es) |
| NES | nuclear export signal |
| NJ | neighbour-joining |
| NLS | nuclear localization signal |
| $OD_{600}$ | optical density at a wavelength of 600 nm |
| ORF | open reading frame |
| PAGE | polyacrylamide gel electrophoresis |
| PBS | phosphate-buffered saline |
| PCNA | proliferating cellular nuclear antigen |
| PCR | polymerase chain reaction |
| $P_i$ | inorganic phosphate |
| PMSF | phenylmethylsulphonyl fluoride |
| pp## | polyprotein ## |
| QRDR | quinolone resistance-determining region |
| RFLP | restriction-fragment length polymorphism |
| RNA | ribonucleic acid |
| RNApol | RNA polymerase |
| RNase | ribonuclease |
| rpm | rotations per minute |
| RPMI | Roswell Park Memorial Institute |
| RVR | right variable region |
| SDS | sodium dodecylsulphate |
| SDS-PAGE | sodium dodecylsulphate-polyacrylamide gel electrophoresis |
| SV40 | simian virus 40 |
| TAE | Tris-acetate-EDTA |
| TBS | Tris-buffered saline |
| TBST | TBS supplemented with 0.1% Tween-20 |
| topoI | type I DNA topoisomerase(s) |
| topoII | type II DNA topoisomerase(s) |
| TOPRIM | topoisomerase-primase |
| UPGMA | unweighted pair group method with arithmetic mean |
| URA | uracil |
| v/v | volume per volume |

| | |
|---|---|
| VOCs | virus orthologous clusters |
| vp## | viral protein ## |
| w/v | weight per volume |
| WHD | winged helix domain |
| YNB | yeast nitrogen base |

# CHAPTER 1

# INTRODUCTION

African swine fever (ASF) is a highly threatening disease for pig husbandry, against which there is neither effective vaccine nor treatment, and its control relies exclusively on sanitary measures. Its etiological agent is a double-stranded DNA arbovirus, the African swine fever virus (ASFV), classified as the only member of the family Asfarviridae (Dixon *et al.*, 2012). ASFV can infect all members of the family Suidae and can be transmitted by direct contact between infected and susceptible animals or indirectly through virus contaminated feed or fomites (EFSA AHAW Panel, 2014). It is also the only DNA virus known to be transmitted by arthropods, in this case soft ticks of the genus *Ornithodoros*, which are of particular relevance for maintaining the endemicity in sub-Saharan countries (Costard, Mur, Lubroth, Sanchez-Vizcaino, & Pfeiffer, 2013). While the virus natural mammalian hosts, the African bushpig and warthog, show no clinical signs upon infection, its introduction in a domestic swine population usually leads to acute forms of disease with nearly 100% mortality, although chronic and unapparent forms of infection may occur, depending on the virus isolate (Blome, Gabriel, & Beer, 2013).

## 1.1 African swine fever

### 1.1.1 Historical synopsis

The first description of African swine fever as a disease distinct from classical swine fever (CSF) came from Kenya, by Montgomery (1921). These findings were later confirmed in South Africa by Steyn in 1928 and 1932 and by De Kock and colleagues in 1940 (Plowright, 1986). These authors showed that warthogs (*Phacochoerus africanus*) and bushpigs (*Potamochoerus larvatus*) comprise a relevant reservoir of the virus and demonstrated that, in domestic pigs, the virus could be harboured up to 10 months and be transmitted up to 7 months after an outbreak, establishing the necessity for a policy of quarantine and complete slaughter to eradicate the infection. Montgomery was also responsible for the clear description of the sub-acute and chronic forms of ASFV, with the latter also being described by Gago da Câmara (1932) in Angola, who in addition proposed an important role for these latter forms of the disease in the transmission cycle of ASFV.

While initially the disease was confined to sub-Saharan countries, in 1957 it was introduced in Europe, through Portugal, probably coming from Angola. Even though this outbreak, the first reported outside of the African continent, was controlled and eradicated, a new ASFV outbreak was reported in Portugal in 1960 (Manso Ribeiro & Azevedo, 1961). From its second introduction ASF became endemic in the Iberian Peninsula, which was only declared ASF-free in 1995 (Arias & Sánchez-Vizcaíno, 2008), and also spread throughout Europe, reaching countries such as France, Italy, Malta, Belgium and the Netherlands, and

also overseas, reaching Cuba (1971 and 1980), the Dominican Republic and Brazil (1978) and Haiti (1979) (Costard *et al.*, 2013; Sánchez-Vizcaíno, Mur, Gomez-Villamandos, & Carrasco, 2015). It was successfully eradicated from all of these countries, with the exception of the Italian island of Sardinia, where ASF remains endemic since 1978. In Africa, ASF has seen a dramatic expansion in the last couple of decades, reaching countries in which the disease had never been reported: Côte d'Ivoire (1996), Nigeria (1997), Togo (1997), Ghana (1999), Burkina Faso (2003) and Chad (2010), as well as the islands of Madagascar (1998) and Mauritius (2007). In 2007 ASF also spread into Eastern Europe, in Georgia *via* the port of Poti, possibly through contaminated fomites imported from the Southeastern region of Africa (Beltrán-Alcrudo, Lubroth, Depner, & De La Rocque, 2008). It quickly spread to neighbouring countries like Armenia, Azerbaijan and the Russian Federation, and at the time of writing it has reached as far as the Ukraine, Poland, Lithuania, Latvia and Estonia (Sánchez-Vizcaíno et al., 2015).

### 1.1.2 Hosts of African swine fever virus

There are several wild members of the family Suidae known to exist in Africa and for which infection with ASFV has been reported: the warthog, the bushpig, the red river hog (*Potamochoerus porcus*) and the giant forest hog (*Hylochoerus meinertzhageni*) (Jori & Bastos, 2009). Due to its area of distribution, in Central Africa and away from areas of domestic pig production, the giant forest hog is thought to have a negligible involvement in the epidemiology of ASF (Jori & Bastos, 2009). Albeit the identification, by Steyn and De Kock, of African warthogs and bushpigs as reservoirs of ASFV, the precise role of the bushpig, and of its relative the red river hog, remains elusive. Their nocturnal habits, non-use of burrows, low population densities and distribution far from areas of historical ASF outbreaks and of warthog distribution reduce their contact with other potential hosts (Jori & Bastos, 2009). Even though ASFV replication in bushpigs and its transmission to pigs and soft ticks has been demonstrated, the mechanisms by which they occur have not been determined (Anderson, Hutchings, Mukarati, & Wilkinson, 1998). Therefore, the warthog is not only considered the original vertebrate host of ASFV but is also thought to be the most important vertebrate reservoir for ASF in Africa (Jori & Bastos, 2009). It participates in a sylvatic cycle of transmission with soft ticks of the genus *Ornithodoros*, to which contribute its widespread geographical distribution and occupation of burrows. While infected adult warthogs show no clinical symptoms of the disease, high viral replication and viremia occur in young animals, usually after being bitten in burrows by infected soft ticks (Thomson, 1985). Since viremia in adult warthogs is low they are unable to infect ticks and so it is during

infection of young warthogs that naive ticks become infected, by feeding on them (Jori & Bastos, 2009).

As previously mentioned, *Ornithodoros* ticks are also important hosts of ASFV. In Africa, ticks of the *O. moubata* complex are a source of infection for both domestic and wild pigs, with pig-to-tick transmission occurring in blood meals (Burrage, 2013). Transstadial, transovarial and sexual transmission between ticks, which has been shown to occur (Plowright, Perry, & Peirce, 1970; Plowright, Perry, & Greig, 1974; Wardley *et al.*, 1983; Rennie, Wilkinson, & Mellor, 2001), as well as persistence of the virus in the ticks for long periods of time allow for the maintenance of ASF in infected areas despite the absence of other infected hosts. Still, despite being well-adapted to the tick host, ASFV can still induce mortality in ticks (Kleiboeker & Scoles, 2001). *O. moubata* ticks are distributed throughout Southern Africa and also in Madagascar, but evidence of their presence in Central Africa is scarce and they are thought to be absent from West Africa (Costard *et al.*, 2013). In Europe there are no *O. moubata* ticks, but *O. erraticus* can be found and have been shown to be an ASFV host and to be associated with the persistence and recurrence of the disease in some regions of Portugal and Spain (Sanchez-Botija, 1963; Caiado, Boinas, & Louza, 1988; Basto *et al.*, 2006; Boinas, Wilson, Hutchings, Martins, & Dixon, 2011). Six other species of *Ornithodoros* ticks have been confirmed to be susceptible to ASFV infection, namely *O. coriaceus*, *O. turicata*, *O. parkeri* and *O. puertoricensis*, from North America and Caribbean, *O. savignyi* from North Africa and *O. sonrai* from West Africa (Vial *et al.*, 2007; Estrada-Pena *et al.*, 2010).

While both African wild suids and *Ornithodoros* ticks are well-adapted to infection by ASFV and survive it with relative ease, in domestic pigs ASF is usually associated with high mortality and high morbidity rates. It seems, however, that in endemic regions the mortality rates have a tendency to lower over time while subclinical or chronic ASFV infections emerge (Sánchez-Vizcaíno, Mur, & Martínez-López, 2012; Costard *et al.*, 2013). These chronically infected animals may end up acting as reservoirs for the disease, contributing to its persistence and to its spread to ASF-free areas, usually through direct contact between animals or indirectly through tick bites or contact with fomites (Sánchez-Vizcaíno *et al.*, 2012). European wild suids like feral pigs and wild boars have also been shown to be susceptible to ASFV, at levels comparable to domestic pigs (Arias & Sánchez-Vizcaíno, 2008). While outbreaks in wild suid populations tend to fade out unless contact with a source of infection, like domestic pigs, is maintained, these animals may facilitate the spread of the disease, especially in areas with free-ranging pigs (Costard *et al.*, 2013).

### 1.1.3 Transmission of African swine fever

Three main cycles of transmission of African swine fever have been described: the sylvatic cycle, the tick-pig cycle and the domestic cycle.

The sylvatic cycle involves transmission between warthogs and *O. moubata* ticks and is well established in the southern and eastern regions of Africa (Jori *et al.*, 2013). Because ASFV infection in adult warthogs is asymptomatic and there is no horizontal or vertical transmission of the virus among these animals, maintenance of infection in the warthog population is dependent on *O. moubata* ticks. These infect young warthogs in infested burrows, allowing subsequent transmission of the virus to naive ticks during blood meals. *O. moubata* colonies are also able to maintain ASFV infection for many months in the absence of blood meals, due to sexual transmission of the virus, contributing to the persistence of the disease in the affected regions. No evidence for the presence of a sylvatic cycle in Western Africa has been found (Jori & Bastos, 2009).

As *Ornithodoros* ticks do not feed exclusively on warthogs and have also been found infesting pig pens in Africa and in the Iberian Peninsula (Haresnape & Mamu, 1986; Caiado *et al.*, 1988; Haresnape, Wilkinson, & Mellor, 1988; Wilkinson, Pegram, Perry, Lemche, & Schels, 1988; Haresnape & Wilkinson, 1989; Oleaga-Pérez, Pérez-Sánchez, & Encinas-Grandes, 1990; Boinas *et al.*, 2011), they easily contribute to the transmission and spread of African swine fever through a tick-pig cycle. Indeed, the occurrence of ASF outbreaks has been found to associate significantly with the presence of *O. erraticus* in Spain (Pérez-Sánchez, Astigarraga, Oleaga-Perez, & Encinas-Grandes, 1994) and the re-emergence of the disease in Portugal in 1999 is thought to be due to the persistence of *O. erraticus* in a previously infected farm (Basto *et al.*, 2006; Boinas *et al.*, 2011).

Once introduced in the domestic pig population, and as these animals usually present high viremia and are frequently maintained in large numbers, ASFV can be transmitted by direct contact between infected and non-infected animals or by contact between pigs and fomites, with no apparent role of soft ticks (Sánchez-Vizcaíno *et al.*, 2012). This constitutes the domestic cycle of ASFV transmission. Movement and/or trade of pigs and the lack of biosecurity measures of control contribute significantly to the spread of ASF in and from endemic regions, as well as the resistance of ASFV to inactivation and its lengthy persistence in pork products (Costard *et al.*, 2013).

### 1.1.4 Molecular epidemiology

While outside of Africa, or in its Western region, the epidemiology of ASF is fairly simple, in East and South Africa it is highly complex, especially due to the simultaneous existence of

sylvatic, pig-tick and domestic pig transmission cycles. Initial studies for genetic characterization of different ASFV isolates were based on restriction-fragment length polymorphism (RFLP) analysis (Wesley & Tuthill, 1984; Dixon & Wilkinson, 1988). However, this method takes very long time to produce results, hindering its usefulness during ASF outbreaks. To overcome this, Bastos *et al.* (2003) took advantage of inter-isolate variation in 415 base pairs of the C-terminal region of ORF B646L, coding for the major capsid protein vp72, to determine relationships between field isolates. This method allowed for the classification of ASFV isolates in ten distinct genotypes, with the largest group, genotype I, being the least diverse and comprising isolates from Europe, South America, the Caribbean and West Africa. On the other hand, the remaining nine genotypes contained isolates from Southern and East Africa, confirming the heterogeneity found by RFLP analysis in isolates from those areas (Blasco, Agüero, Almendral, & Viñuela, 1989). Further studies on vp72 genotyping (Lubisi, Bastos, Dwarka, & Vosloo, 2005; Boshoff, Bastos, Gerber, & Vosloo, 2007) extended the number of genotypes for which ASFV isolates could be classified to twenty-two. All of these studies hinted for a correlation between the diversity observed for Southern and Eastern African isolates and the existence of a sylvatic cycle in those areas, and confirmed that most isolates from outside of Africa are contained in genotype I.

To obtain higher resolution of the relationships between very similar isolates Gallardo *et al.* (2009) added further layers of genotyping to the vp72 method, by fully sequencing ORF E183L, coding for the late-expressed structural protein vp54, essential for the assembly of ASFV particles (Rodríguez *et al.*, 2004), as well as sequencing of a central, highly variable region of ORF B602L, which codes for a possible molecular chaperone involved in the folding of vp72 and required for the correct assembly of the viral capsid (Epifano, Krijnse-Locker, Salas, Rodríguez, & Salas, 2006a). As desired, this improvement of the method allowed the discrimination of isolates previously grouping in the same genotype. Genotyping of vp54 originated four sub-clusters (a, b, c and d) within genotype I, as well as two sub-clusters (a and b) for genotypes V, X and XX. In addition, genotyping of B602L permitted the discrimination between two groups of very similar Kenyan isolates belonging to genotype IX.

With the fast development of sequencing technologies, and even though ASFV has a relatively large DNA genome, sequencing of the entire ASFV genome is relatively fast, though probably not feasible in the event of an outbreak, either due to time or technological constraints. Nevertheless, sixteen ASFV isolates have had their genome sequenced (Yáñez *et al.*, 1995; Chapman, Tcherepanov, Upton, & Dixon, 2008; de Villiers *et al.*, 2010; Bishop *et al.*, 2015; Portugal *et al.*, 2015) and in a recent whole-genome phylogenetic analysis (Portugal

*et al.*, 2015) the distribution of fourteen of those isolates was found to overlap with what was found using the vp72 method, with high bootstrap support. This shows that, with some technological improvement, whole-genome genotyping may elucidate even further the relationships between circulating ASFV isolates.

As mentioned, and until recently, only isolates of vp72 genotype I were found outside of Africa. However, the recent introduction of ASF in Georgia changed this scenario, as initial studies showed that the circulating virus bared genetic resemblance with isolates from Madagascar and Mozambique (Rowlands *et al.*, 2008), which are of genotype II.

### 1.1.5 Morphology of ASFV

ASFV is a large, enveloped virus, with a particle of 200 nm with icosahedral morphology and composed by several concentric domains: an outer envelope, the viral capsid, an inner envelope, the core shell and the nucleoid (Figure 1).

The outer envelope is only present in extracellular virions that undergo budding, being acquired from the plasma membrane of the host cell (Breese & DeBoer, 1966). It does not exist in mature viral particles released due to lysis of the infected cell (Breese & DeBoer, 1966), even though these are also infectious. In order to egress, particles formed in the viral factory bind to microtubules and, *via* a microtubule-dependent and kinesin-mediated mechanism (Matos & Carvalho, 1993; Jouvenet, Monaghan, Way, & Wileman, 2004), are transported to the budding site at the cell surface. The viral protein vp12 is reported to be present in the external envelope (Carrascosa, Saastre, Gonzalez, & Viñuela, 1993). A viral CD2 homolog, pEP402R, is also thought to be present (Salas & Andrés, 2013) in the outer envelope due to its ability to promote hemadsorption (Rodríguez, Yáñez, Almazán, Viñuela, & Rodriguez, 1993).

The viral capsid is formed by 1892 - 2172 capsomers, each 13 nm in diameter with the appearance of an hexagonal prism with a central hole (Carrascosa *et al.*, 1984), with the major capsid protein being vp72 (Tabarés, Martinez, Gonzalvo, & Sánchez-Botija, 1980; Esteves, Marques, & Costa, 1986). Its assembly occurs on the convex face of viral membranes (García-Escudero, Andrés, Almazán, & Viñuela, 1998). Although not being a structural protein, pB602L acts as a chaperone promoting the correct folding of vp72 (Cobbold, Windsor, & Wileman, 2001) and is essential for the correct assembly of the viral capsid (Epifano *et al.*, 2006a). Protein pE120R is also a component of the capsid and has been shown to be involved in the transport of mature ASFV particles from the viral factory to budding sites at the plasma membrane (Andrés *et al.*, 2001b), probably due to its interaction with the microtubules or microtubule-associated proteins. Protein pB438L has been proposed to be an

integral part of the capsid due to its importance for the correct acquisition of the icosahedral structure (Epifano, Krijnse-Locker, Salas, Salas, & Rodríguez, 2006b).

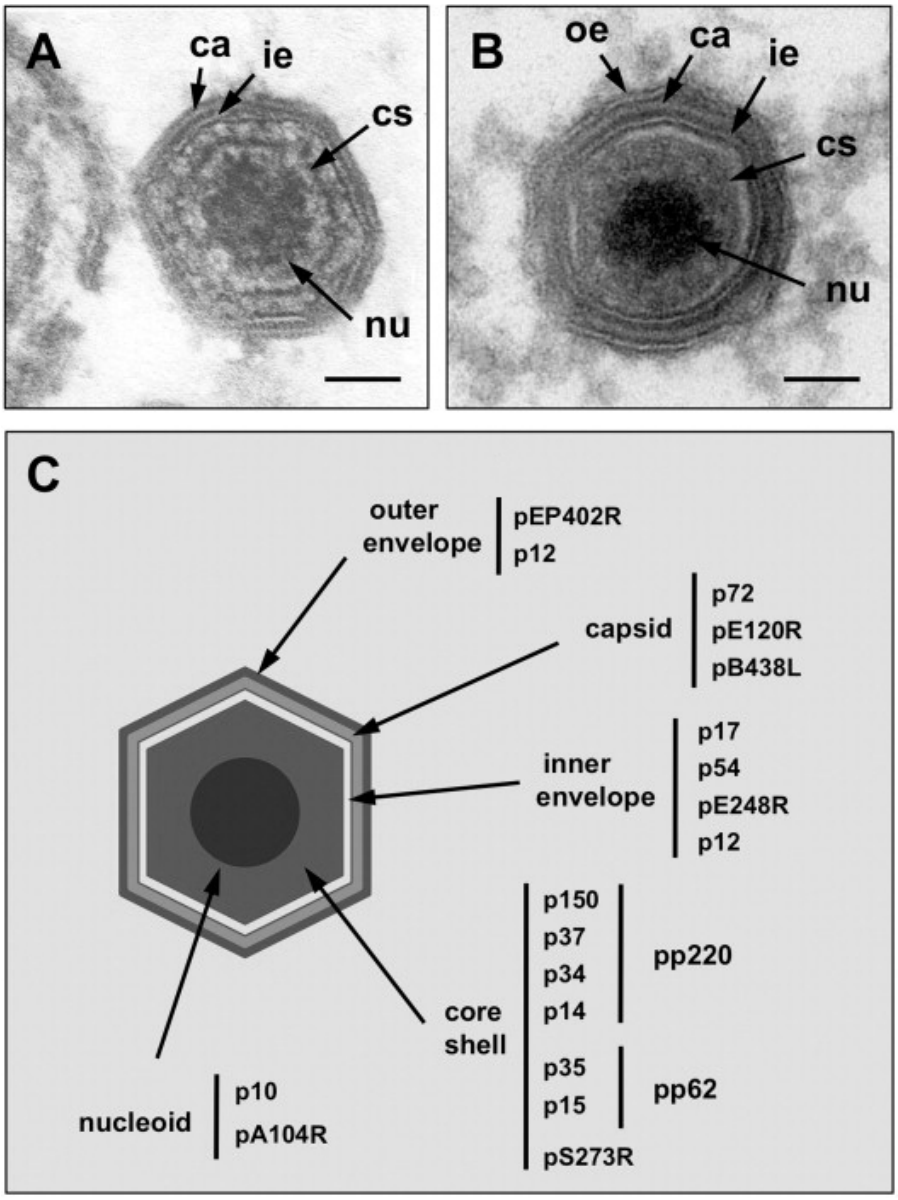


**Figure 1.** Structure and protein composition of ASFV particle.
(A) Electron micrograph of an intracellular full ASFV particle and (B) of an extracellular mature ASF virion. For both panels, the different structural domains are indicated: outer envelope (oe), capsid (ca), inner envelope (ie), core shell (cs) and nucleoid (nu). (C) Illustration of the localization of ASFV structural proteins. Figure and legend adapted from Salas and Andrés (2013).

A clearly visible morphological sign of viral assembly is the accumulation of membranes within the viral factory. These membranes will act as precursors for the formation of the inner envelope of the viral particle. It was recently shown that this accumulation of membranes is a result of the rupture of endoplasmic reticulum (ER) membranes, inducing an accumulation of open membranes in the cytoplasm which are recruited (by an unknown mechanism) to the

viral factory (Suarez *et al.*, 2015). This phenomenon was already suspected (Andrés, García-Escudero, Simón-Mateo, & Viñuela, 1998) and is supported by the co-localization of several ER proteins to membranes in viral factories, in mature virions and in extracellular viral particles (Rouiller, Brookes, Hyatt, Windsor, & Wileman, 1998; Suarez *et al.*, 2015). No continuities between the ER membranes and viral membranes were detected, suggesting that there is indeed a rupture of the ER and not just a modification/reorganization. As the membrane precursors assemble into the inner envelope, their convex side becomes coated with capsid proteins, namely vp72, shaping them into icosahedral morphology, while the concave side is associated with core proteins. ER rupture was shown to be dependent of late protein synthesis (Suarez *et al.*, 2015) and, in the absence of protein vp17, previously shown to localize to the inner envelope (Suárez, Gutiérrez-Berzal, Andrés, Salas, & Rodríguez, 2010), viral membrane precursors failed to form icosahedrals and to become coated by vp72, hinting for an important role of vp17 in this step. Curiously, protein vp12, which was found in the outer envelope of the virion, is also detected in the inner envelope (Salas & Andrés, 2013). This dual distribution may be a consequence of vp12 being a transmembrane protein (Alcamí *et al.*, 1992), of the inner envelope apparently being derived from ER membranes, and of the ER also being part of the pathway that directs proteins to the plasma membrane (Zanetti, Pahuja, Studer, Shim, & Schekman, 2012), from which the outer envelope is derived during budding. Also present in the inner envelope is viral protein pE248R (Rodríguez, Nogal, Redrejo-Rodríguez, Bustos, & Salas, 2009).

The ASFV core shell is constituted mainly by proteins produced from the proteolytic processing of viral polyproteins pp220 and pp62 (Andrés, Alejo, Salas, & Salas, 2002), namely vp150, vp37, vp34 and vp14, derived from pp220 (Simón-Mateo, Andrés, & Viñuela, 1993; Andrés, Simón-Mateo, & Viñuela, 1997), and vp35 and vp15, from pp62 (Simón-Mateo, Andrés, Almazán, & Viñuela, 1997). The protease responsible for this processing, viral protein pS273R, is also present in the core shell (Andrés, Alejo, Simón-Mateo, & Salas, 2001a), suggesting a connection between proteolytic processing and core assembly. Formation of the core shell occurs simultaneously to assembly of the capsid and of the inner envelope, and it is thought to rely on transport of core proteins through serpentine-like structures formed by precursor membranes of the inner envelope (Suarez *et al.*, 2015).

The nucleoid is composed by the viral genome and DNA-binding proteins. Of these, pA104R, a histone-like protein, has been shown to bind single-stranded and double-stranded DNA and was reported to localize in the nucleoid (Borca *et al.*, 1996). vp10 has also been shown to be a DNA-binding protein and a constituent of the virion (Muñoz, Freije, Salas,

Viñuela, & López-Otín, 1993), and is thus thought to localize to the nucleoid. The nucleoid is thought to be the last structure to be formed during ASFV morphogenesis, since icosahedral particles devoid of nucleoid are often observed in viral factories. The accepted model for nucleoid formation (Brookes, Hyatt, Wise, & Parkhouse, 1998) suggests that the viral DNA is first partially condensed, probably with the help of DNA-binding proteins like pA104R and near membranes of the viral factory, and is then inserted, at a single vertex, into the empty viral particle. Further maturation of the virion involves sealing of the capsid opening and further condensation of the nucleoid material.

### 1.1.6 Replication and transcription of ASFV

Due to the characteristics of its genome and to its replication mostly in the cytoplasm of the host cells ASFV is classified as a nucleo-cytoplasmic large DNA virus (NCLDV), a monophyletic group that also encompasses the viral families Poxviridae, Iridoviridae, Ascoviridae, Phycodnaviridae, Mimiviridae, Marseilleviridae, Megaviridae, Pandoraviridae and Pithoviridae. Recently, it was proposed that all these families should be classified under the order Megavirales (Colson *et al.*, 2013). As previously stated, the genome of African swine fever virus is composed by a double-stranded DNA molecule, of 170 to 190 kilobase pairs (kbp) depending on the isolate (Portugal *et al.*, 2015). Its termini contain inverted repeats and covalently closed ends forming 37 nucleotide-long hairpin loops (González, Talavera, Almendral, & Viñuela, 1986), similar to those found for vaccinia virus of family Poxviridae (Baroudy, Venkatesan, & Moss, 1982). The ASFV genome is composed of a highly conserved central region (CCR) flanked by two variable regions, one on the left (LVR) and the other on the right (RVR) of the CCR (Chapman *et al.*, 2008). It is predicted to code for 151 to 167 open reading frames (ORFs), again depending on the isolate, with most of the variation in terms of gene number and length deriving from multigene families – MGF 100, 110, 300, 360 and 505/530 – present in the LVR and RVR. While most ORFs were named based on *Eco*RI restriction enzyme fragmentation, gene orientation left (L) or right (R), and number of amino acids encoded (example: P1192R), the MGFs have been named according to the average number of amino acids encoded by their respective ORFs, the direction in which they are read and the position from the left genome end. The first complete ASFV sequence was that of the Vero-adapted isolate Badajoz71 (Ba71V) (Yáñez *et al.*, 1995), and since then fifteen other isolates have been fully sequenced (Chapman *et al.*, 2008; de Villiers *et al.*, 2010; Bishop *et al.*, 2015; Portugal *et al.*, 2015).

The genes encoded by the ASFV genome are present in both DNA strands and no bias for either strand has been observed. Still, in some regions of the genome neighbouring genes are

found in the same orientation, especially in MGFs (which are thought to be a product of gene duplication). Genes are closely spaced and upstream of each gene there is a short sequence containing the viral promoter, recognized by the viral-encoded RNA polymerase. According to the temporal regulation of their expression, ASFV genes are classified as *immediately early*, *early*, *intermediate* and *late* (Rodríguez & Salas, 2013) (Figure 2). Expression of the

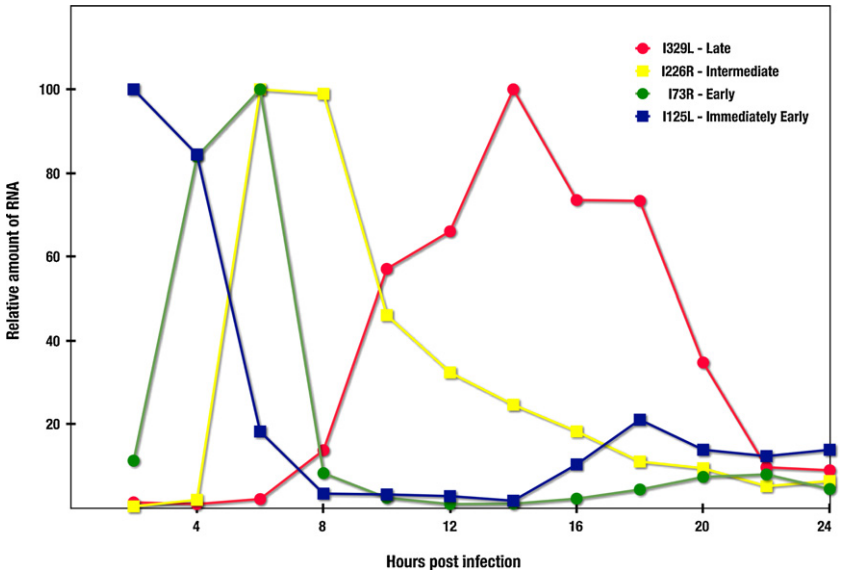


**Figure 2.** Accumulation kinetics for immediately early (blue, I215L), early (green, I73R), intermediate (yellow, I226R) and late (red, I329L) transcripts throughout ASFV infection.
Steady state RNA levels for the different transcripts were detected and measured using primer extension assays. The quantity of each transcript is plotted as the percentage of the maximum level. Figure and legend adapted from Rodríguez and Salas (2013).

first two classes of genes starts prior to viral DNA replication, but the immediate early are repressed first through a process that requires synthesis of viral proteins. It is likely that the immediate early genes are involved in the regulation of host cell responses to infection, and this is supported by the fact that I215L, an ubiquitin-conjugating E2 enzyme that may be involved in targeting host proteins for degradation, was found to be an immediate early gene. On the other hand, the onset of expression of intermediate and late genes requires the start of viral DNA replication, and their expression is not detected in infected cells treated with an inhibitor of DNA replication, like cytosine arabinoside (AraC). Transcription of intermediate genes starts before the beginning of late gene expression and their products are often required for synthesis of late viral proteins. In accordance, the start of late gene expression, at 6 - 8 hours post infection, is more-less coincident with the peak of expression of intermediate genes.

Even though there is no clear consensus on the subject, replication of the ASFV genome is frequently thought to occur in a similar fashion to that of poxviruses, namely vaccinia virus.

In fact, the genomes of these viruses have several common characteristics, like cytoplasmic replication, covalently-linked DNA ends forming hairpin loops, formation of replication intermediates consisting of head-to-head and tail-to-tail concatemers (Baroudy, Venkatesan, & Moss, 1983; González *et al.*, 1986; Caeiro, Meireles, Ribeiro, & Costa, 1990; Rojo, García-Beato, Viñuela, Salas, & Salas, 1999) and genes encoding several proteins involved, or at least predicted to be involved, in viral DNA replication (Dixon, Chapman, Netherton, & Upton, 2013; Moss, 2013). Thus, according to the proposed self-priming model for ASFV replication (Figure 3), its start is dependent on the introduction of a single-strand nick near the genome termini. The nick exposes a 3' OH group to which deoxynucleotides can be added by DNA polymerase. Because of the inverted repeats in the genome termini, which generate self-complementarity in nascent and template strands, these strands fold back to form a self-priming hairpin structure. Progression of the replication complex results in the formation of a concatemer, which is resolved to individual units with cross-linked termini probably around the time of genome encapsidation. While the bulk of ASFV DNA replication takes place in perinuclear cytoplasmic viral factories (Breese & DeBoer, 1966), ASFV requires the host cell nucleus for proper development of infection (Ortin & Viñuela, 1977) and an initial stage of replication inside the nucleus has been reported (Tabarés & Sánchez Botija, 1979; García-Beato, Salas, Viñuela, & Salas, 1992b; Rojo *et al.*, 1999; Ballester *et al.*, 2010). This contrasts with what is observed for poxviruses, which replicate exclusively in the cytoplasm even though they recruit several cellular nucleus-derived factors. However, the role of the nuclear step of DNA replication in ASFV infection is still unknown and under debate.

Of the many ORFs encoded by its genome, ASFV is predicted or known to code for many proteins involved in DNA replication and repair and also in transcription (Dixon *et al.*, 2013; Rodríguez & Salas, 2013). Since the virus cycle is mainly cytoplasmic, and there seems to be no recruitment of the main players of the host cell replicative and/or transcriptional machinery to the viral factory, ASFV is required to code for all the proteins essential for its infection. Among these are an α-like DNA polymerase (G1211R), a putative primase (C962R), a PCNA homolog (E301R), an error prone DNA polymerase X (O174L) and a DNA ligase (NP419L), involved in DNA repair and essential for infection of swine macrophages, quite a few DNA and RNA helicases, several subunits of an RNA polymerase complex and transcription factors and even a type II DNA topoisomerase (P1192R). Indeed, considering that ASFV has a rather large DNA genome with terminal inverted repeats and covalently closed ends, a type II topoisomerase may be required for the resolution of topological problems arising from viral DNA replication and/or transcriptional events.

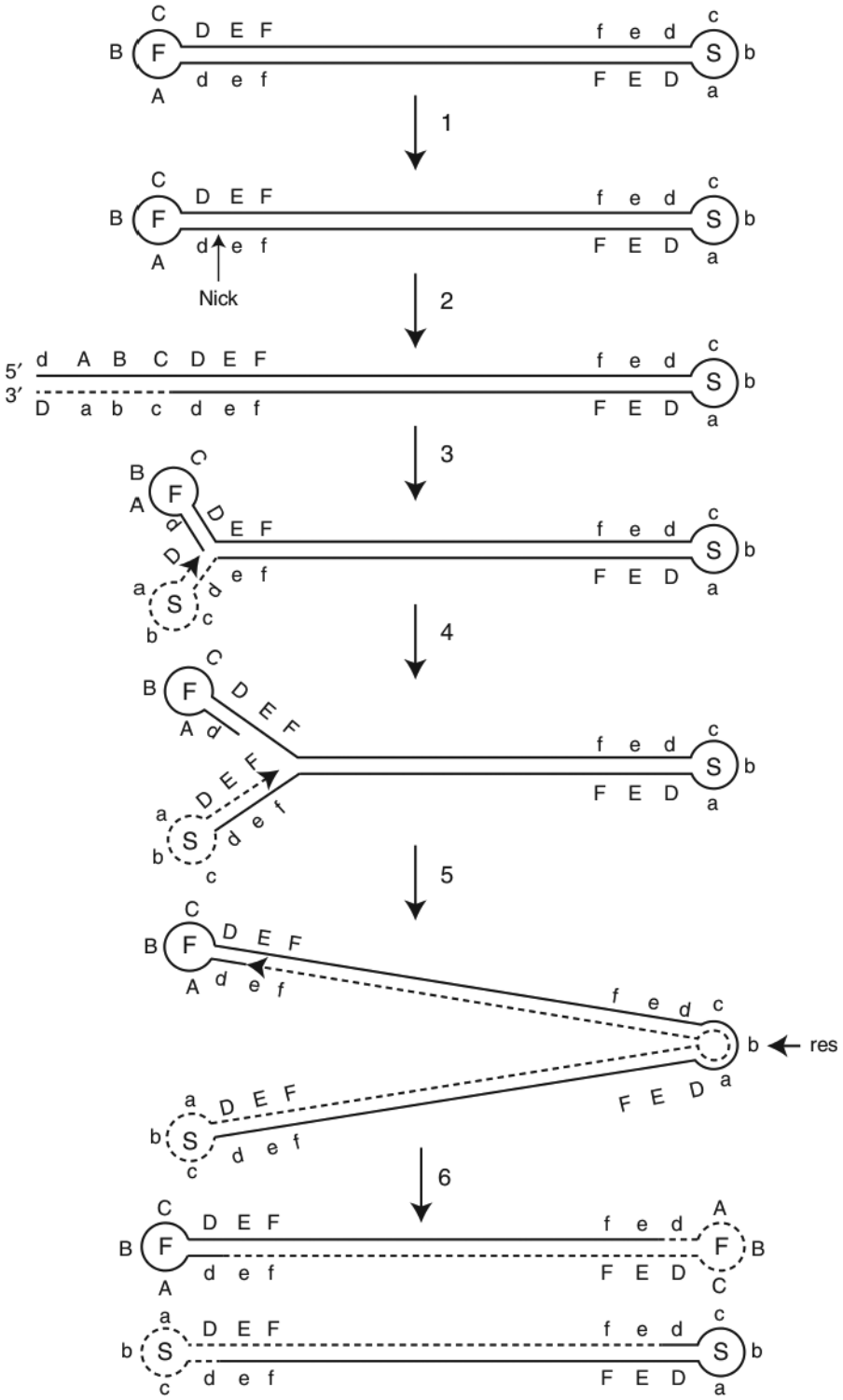


**Figure 3.** Schematic representation of the self-priming model for replication of the ASFV genome, adapted from the model proposed for poxviruses.

Letters F and S within the hairpin loops represent the fast and slow electrophoretic mobilities of DNA fragments containing inverted and complementary hairpin sequences. Dashed lines represent newly synthesized DNA with arrowheads at the 3' OH ends. Complementary sequences are depicted by upper- and lowercase letters. The site of resolution (res) within the concatemer junction is indicated. Figure and legend adapted from Moss (2013).

## 1.2 DNA topoisomerases

### 1.2.1 Brief considerations on the discussion of DNA topoisomerases and topological problems

DNA topoisomerases are ubiquitous enzymes, present in all domains of life. Still, explanation of their activity and of why they are required is simplest in the context of a closed circular DNA molecule. Thus, for the sake of clarity, reactions catalysed by these enzymes will often be described for DNA rings, as in prokaryotic organisms, even though they also occur in linear chromosomes, as in eukaryotes (Wang, 2002; Schoeffler & Berger, 2008; Nitiss, 2009a). This is because long linear molecules are often organized into loop-containing structures, or domains, and since the break in the DNA is usually too far away from sites in which topological problems occur, changes in winding cannot occur with reasonable kinetics without the intervention of DNA topoisomerases. Additionally, the topological consequences of an advancing fork, be it during replication or transcription, and the roles of different DNA topoisomerases, depend of the ability of the respective protein complexes to rotate around the DNA. Still, these topological constraints are thought to be limited to closed topological domains, restricting the diffusion of torsional changes to larger chromosomal regions (Kanaar & Cozzarelli, 1992; Postow, Crisona, Peter, Hardy, & Cozzarelli, 2001; Postow, Hardy, Arsuaga, & Cozzarelli, 2004; Fritsche, Li, Heermann, & Wiggins, 2012)

### 1.2.2 Overview of DNA topoisomerases

DNA topoisomerases are enzymes that modulate the topological state of DNA molecules. They are essential in processes such as DNA replication, recombination, repair and transcription, that cause overwinding and/or underwinding of the DNA which, if not resolved, may compromise genomic stability and cellular viability. All DNA topoisomerases act by creating breaks in DNA through the nucleophilic attack of a tyrosine catalytic residue on the phosphodiester backbone of the DNA molecule, generating a covalent phosphotyrosine link (Schoeffler & Berger, 2008; Pommier, Leo, Zhang, & Marchand, 2010). They are classified into two types, on the basis of their strand scission activity: type I topoisomerases (topoI), which are ATP-independent enzymes that produce transient single-stranded breaks in the DNA, thereby facilitating unwinding and changing the topological linking number by steps of one, and type II topoisomerases (topoII), that require ATP in order to transiently cleave both strands of DNA which, followed by double-stranded DNA passage, changes the linking number by steps of two. Each type is also subdivided into several classes according to structural and functional similarities. Thus, type I topoisomerases are subdivided into topoIA,

topoIB and topoIC, while type II topoisomerases are subdivided into topoIIA and topoIIB (Forterre, Gribaldo, Gadelle, & Serre, 2007; Schoeffler & Berger, 2008).

Type IA and IB differ in the mechanism by which they act and in the way they establish the single-stranded break in the DNA. TopoIA enzymes generate phosphotyrosine links with the free 5' end of the DNA break, originating a topoisomerase-bridged single-stranded gap through which a second DNA strand, or even a double-stranded segment, is passed (Brown & Cozzarelli, 1981). This activity is dependent on divalent ions (namely $Mg^{2+}$) (Domanico & Tse-Dinh, 1991) and allows topoIA to relax negative supercoils as well as unknotting and decatenating DNA regions containing nicks or in single-stranded form (Brown & Cozzarelli, 1981). TopoIA examples are bacterial and archaeal topoisomerase I and topoisomerase III, eukaryotic topoisomerase III and the mimivirus topoisomerase IA (Forterre *et al.*, 2007). While topoisomerase I is specialized for relaxing negative supercoils, topoisomerase III is primarily implicated in the resolution of recombination intermediates, both in bacteria and in eukaryotes (Wang, 2002; Schoeffler & Berger, 2008). Also included in the topoIA is the reverse gyrase, a topoisomerase present in all hyperthermophilic archaea and bacteria (Forterre *et al.*, 2007). In addition to its ability of relaxing supercoiled DNA, reverse gyrase can also introduce positive supercoils in DNA molecules using ATP as a driving force. Positive supercoiling is thought to be a mechanism by which hyperthermophilic organisms prevent temperature-induced DNA relaxation (Forterre, Bergerat, & Lopex-Garcia, 1996).

In opposition to topoIA, topoIB enzymes create a nick in one DNA strand by establishing a phosphotyrosine link with the free 3' end (Champoux, 1981). This allows the free 5' OH end to rotate around the intact strand, with the driving force of this rotation being the energy stored in DNA supercoils (Schoeffler & Berger, 2008). This mechanism allows topoIB enzymes to relax both positive and negative supercoils in an ATP- and divalent ion-independent manner (Dynan, Jendrisak, Hager, & Burgess, 1981). Type IB topoisomerases include the eukaryotic topoisomerase I, which is ubiquitous and responsible for the major topoisomerase I activity in eukaryotic organisms, topoisomerase I from poxviruses, and topoisomerase IB from some bacteria and from mimivirus (Forterre *et al.*, 2007). Moreover, in vertebrates two topoIB are present, with one localizing to the nucleus and the other to mitochondria (Zhang *et al.*, 2001), and it is speculated that they result from a gene duplication event (Zhang, Meng, Zimonjic, Popescu, & Pommier, 2004). TopoIB enzymes present structural and functional similarities with tyrosine recombinases (Cheng, Kussie, Pavletich, & Shuman, 1998), hinting for a common origin.

TopoIC has a single member, topoisomerase V, found in archaeas of the genus *Methanopyrus* (Taneja, Patel, Slesarev, & Mondragón, 2006). It relaxes both positively and

negatively supercoiled DNA, in the absence of divalent ions, and like type IB enzymes it establishes a phosphotyrosine bond with the 3'end of the cleaved DNA strand. However, it does so by a mechanism distinct of that of type IB topoisomerases (Taneja *et al.*, 2006).

Both type IIA and type IIB topoisomerases form an enzyme-bridged double-stranded break, through which they pass another DNA duplex. Nevertheless, they differ in their structural organization and phylogenetic distribution. Type IIB enzymes are heterotetrameric ($A_2B_2$) complexes that generate 5'-phosphotyrosine links to DNA and require ATP and $Mg^{2+}$ to function properly (Bergerat, Gadelle, & Forterre, 1994). They are capable of relaxing both negative and positive supercoils in the DNA, as well as performing decatenation. The sole member of type IIB, topoisomerase VI, is ubiquitous in archaea and is also present in plants, green and red algae and some species of bacteria (Forterre *et al.*, 2007). The A-subunit of topoisomerase VI, which contains the catalytic tyrosine, bears resemblance with the eukaryotic Spo11 protein, which was shown to participate in meiotic recombination (Forterre *et al.*, 2007; Schoeffler & Berger, 2008).

Type IIA topoisomerases have the ability to relax supercoils, to resolve knots and tangles in the DNA, and allow for the decatenation and proper segregation of intertwined DNA molecules following replication, in an ATP-dependent manner. In lower eukaryotes, like *Saccharomyces cerevisiae* and *Drosophila melanogaster*, type IIA topoisomerases are represented by a single isoform (Goto & Wang, 1984; Nolan, Lee, Wyckoff, & Hsieh, 1986), while higher eukaryotes and bacteria encode two forms of these enzymes. DNA gyrase (Gellert, Mizuuchi, O'Dea, & Nash, 1976a) and topoisomerase IV (Kato *et al.*, 1990) are closely related enzymes that can be found in bacteria, but with distinct physiological roles. Both are heterotetrameric, with DNA gyrase being composed by subunits GyrB and GyrA ($B_2A_2$), while topoisomerase IV, or topoIV, is built of subunits ParE and ParC ($E_2C_2$), with respect to their assembly order and homology (Gadelle, Filée, Buhler, & Forterre, 2003). DNA gyrase has the unique ability of introducing negative supercoils into positively supercoiled or relaxed DNA (Gellert *et al.*, 1976a), owing this ability to its highly-conserved (among DNA gyrases) C-terminal domain (present in the C-terminus of the GyrA subunit) (Kramlinger & Hiasa, 2006). This special activity of DNA gyrase is of importance in maintaining the high negative superhelicity found in many bacterial genomes (Drlica, 1992). Aside from introducing negative supercoils, DNA gyrase is also capable of removing negative or positive supercoils from DNA molecules and of promoting decatenation (Cozzarelli, 1980). Aside from its supercoiling function, gyrase is of importance for the initiation of DNA replication and transcription (Pommier *et al.*, 2010). TopoIV is capable of performing the same reactions as DNA gyrase, with exception of DNA supercoiling (Kato *et al.*, 1990). In

addition, it has a higher affinity for relaxing positively supercoiled DNA and for decatenating plasmid DNA (Peng & Marians, 1993), which explains its preeminent role in chromosome separation during DNA replication (Wang, Reyes-Lamothe, & Sherratt, 2008).

In vertebrates, two highly identical type II topoisomerases also exist, topoisomerase IIα and IIβ (Gadelle *et al.*, 2003). They are extensively homologous for most of their sequence, with exception for a degenerate C-terminal region containing nuclear localization signals (Mirski *et al.*, 1997; Mirski, Gerlach, & Cole, 1999) and regulatory sites (Isaacs *et al.*, 1998). While their mechanism of action seems to be equal, they diverge in their expression pattern throughout the cell cycle and in their tissue distribution, suggesting that their cellular roles are different (Nitiss, 2009a). TopoIIα is highly expressed in actively-dividing cells and is essential in all cells (Grue *et al.*, 1998; Carpenter & Porter, 2004). It relaxes positive supercoils more efficiently than it does negative supercoils (McClendon, Rodriguez, & Osheroff, 2005) and is thought to be the major decatenase in vertebrates (Nitiss, 2009a), thereby having a pivotal role in DNA replication and chromosome segregation. On the other hand, topoIIβ does not display preference for positively or negatively supercoiled DNA, a property shared with topoII from organisms containing a single isoform (like yeast and *Drosophila*) (Osheroff, Shelton, & Brutlag, 1983; Roca & Wang, 1996). It is non-essential (Grue *et al.*, 1998) and mostly expressed in quiescent cells (Pommier *et al.*, 2010), having a critical role in transcription (Ju *et al.*, 2006). Given the sequence similarities between topoIIα and topoIIβ, the substrate specificity of the former is thought to be conferred by the variable C-terminal region. In accordance, when the C-terminal domains of both enzymes are swapped, the topoIIβ core is able to support cell proliferation in the absence of native topoIIα (Linka *et al.*, 2007).

In viruses, type IIA topoisomerases have been identified in several NCLDV families, such as Iridoviridae, Asfarviridae, Mimiviridae and Phycodnaviridae (Iyer, Aravind, & Koonin, 2001; Yutin, Colson, Raoult, & Koonin, 2013), as well as in bacteriophages of the T4 superfamily (Kreuzer, 1998).

For the purpose of this work, type IIA DNA topoisomerases will be henceforth referred simply as type II DNA topoisomerases, or topoII, while topoisomerase VI, although being a type II topoisomerase but of the IIB family, will be referred simply as topoisomerase VI, or topoVI.

A schematic summary of known topoisomerases and their actions on DNA topology is shown on Figure 4.

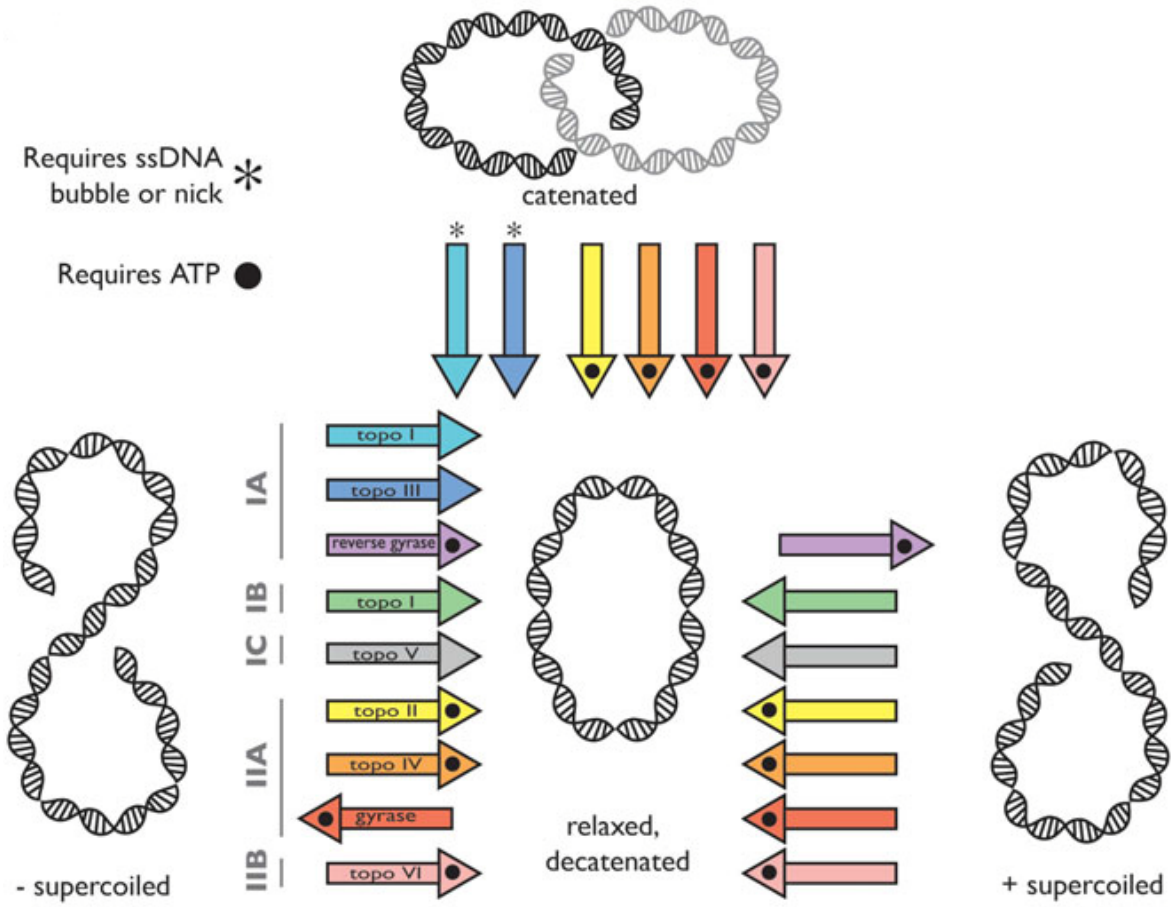


**Figure 4.** Roles of DNA topoisomerases in maintaining cellular DNA topology.
Members of five DNA topoisomerase subfamilies are shown as arrows, labelled on the left, indicating their activities. The represented activities are of relaxation of negative (from left to center) and positive (from right to center) supercoils, of decatenation of linked DNA (from top to center) and of introduction of positive (from center to right) or negative (from center to left) supercoils. Requirement of ATP or of the presence of a single-stranded DNA gap or nick for a particular activity are also indicated. Figure and legend adapted from Schoeffler and Berger (2008).

### 1.2.3 The catalytic reaction of type II DNA topoisomerases

While type II topoisomerases from eukaryotes and from eukaryotic viruses are homodimeric enzymes, bacterial topoIIs are heterotetramers and there is even the case of the bacteriophage T4 type II DNA topoisomerase which is a heterohexamer ($A_2B_2C_2$) (Schoeffler & Berger, 2008). Even so, type II topoisomerases are always present in dyad symmetry, using ATP (Gellert, O'Dea, Itoh, & Tomizawa, 1976b) to generate a transient double-stranded DNA break, with each strand being cleaved by half of the topoII complex. The double-stranded break is not blunt, but instead is composed of 5' protruding ends that are four-nucleotide long, as well as 3' recessed ends with free hydroxyl groups (Morrison & Cozzarelli, 1979; Liu, Rowe, Yang, Tewey, & Chen, 1983; Sander & Hsieh, 1983). By action of the topoII an intact

DNA strand is passed through the transient break, which is then resealed (Liu, Liu, & Alberts, 1980). While a reaction cycle of a type II topoisomerase requires two ATP molecules, the hydrolysis of these molecules is not synchronized and occurs at different steps of the cycle (Harkins & Lindsley, 1998; Harkins, Lewis, & Lindsley, 1998). An important feature of the DNA cleavage reaction of topoII is the phosphotyrosine link that is formed between a tyrosine at the catalytic site of each enzyme subunit and the phosphodiester backbone of the DNA molecule (Tse, Kirkegaard, & Wang, 1980). This bond implies that the strand breaks are covalently bound to the topoisomerase subunits, forming a covalent complex (Morrison & Cozzarelli, 1979; Liu *et al.*, 1983; Sander & Hsieh, 1983; Osheroff & Zechiedrich, 1987), and are thus protected from spurious rearrangements or recombination. Establishment of the phosphotyrosine bond requires the collaboration of other residues that compose the active site, including three acidic residues involved in binding a divalent cation (Berger, Gamblin, Harrison, & Wang, 1996; Dong & Berger, 2007), namely $Mg^{2+}$, which is absolutely required for the reaction (Osheroff, 1987). Reversal of the cleavage reaction, and thus of the phosphodiester bond, can be accomplished without hydrolysis of ATP as the energy contained in the phosphodiester bond is conserved in the phosphotyrosine bond.

The structure of eukaryotic type II topoisomerases is highly conserved and most structural studies have been based on the yeast topoII model (Berger *et al.*, 1996; Fass, Bogden, & Berger, 1999; Classen, Olland, & Berger, 2003; Dong & Berger, 2007; Schmidt, Osheroff, & Berger, 2012). Still, their overall organization is similar to that of the bacterial DNA gyrase and topoisomerase IV (Forterre *et al.*, 2007; Schoeffler & Berger, 2008). Eukaryotic type II topoisomerases are typically composed of an N-terminal ATPase domain (Lindsley & Wang, 1991), a central catalytic region containing the TOPRIM domain (Aravind, Leipe, & Koonin, 1998) and the breakage-reunion domain (Berger *et al.*, 1996) which contains the catalytic tyrosine residue, and a less-conserved C-terminal regulatory domain (Champoux, 2001) (Figure 5).

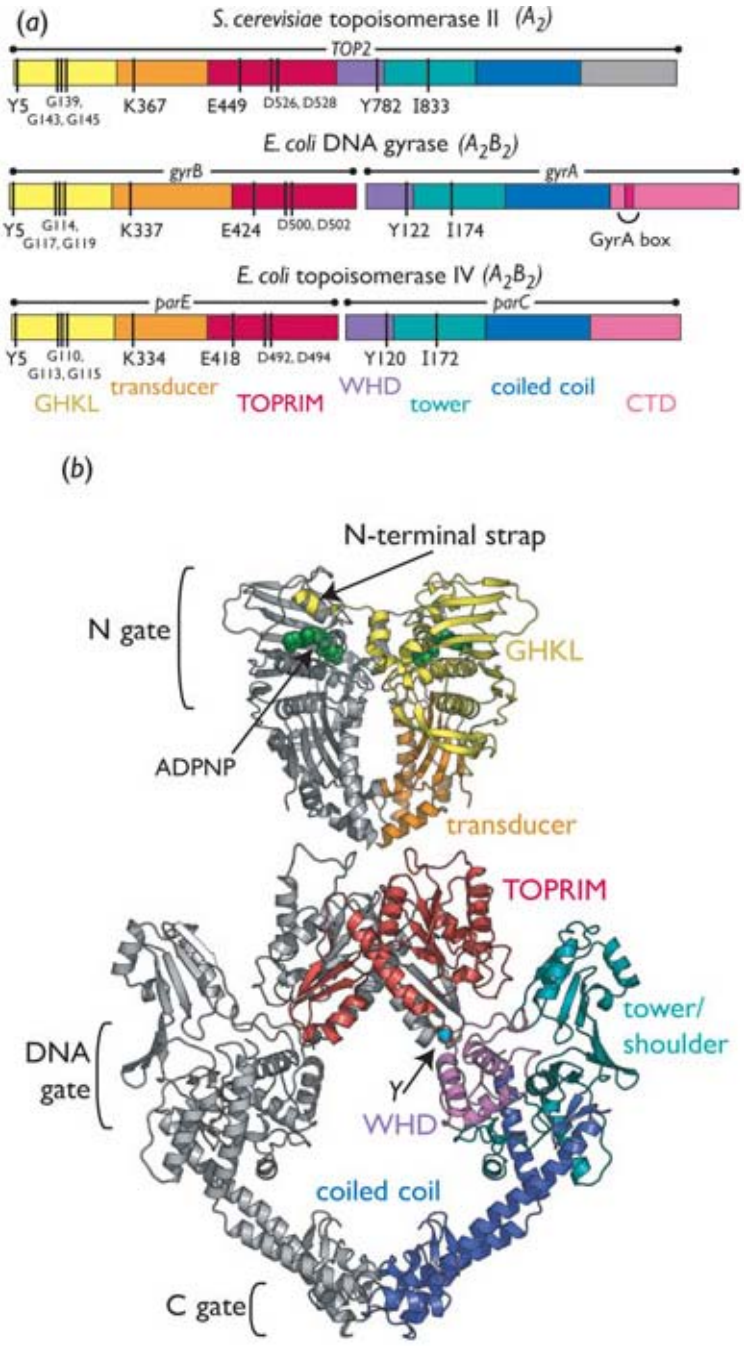


**Figure 5.** Structure of type II DNA topoisomerases.
(*a*) Domain structure of representative eukaryotic and prokaryotic type II topoisomerases. Domains are indicated by colour and important residues by black bars. Gene names are shown above each primary structure along with the subunit compositions. (*b*) Structure of *S. cerevisiae* Top2p based on structures for the ATPase domain of Top2p and the breakage-reunion domain of *E.coli* DNA gyrase. For clarity, one subunit of each dimer is shown in grey. Colours of the legend correspond to the colours of each domain in the structure. The catalytic tyrosine is indicated by a Y, and gate regions are indicated to the left. Figure and legend adapted from Schoeffler and Berger (2008).

As topoII are symmetric enzymes, during the reaction cycle several dimerization interfaces, or gates, of each monomer associate and dissociate in a controlled fashion, facilitating the passage of one DNA duplex through another (Roca & Wang, 1992; Berger *et al.*, 1996; Roca, Berger, Harrison, & Wang, 1996; Morais Cabral *et al.*, 1997). This is the case for the ATP-binding domain, also termed as the N gate or the ATP gate, as collaboration between the subunits couples binding of ATP to formation of the topoisomerase dimer (Wigley, Davies, Dodson, Maxwell, & Dodson, 1991; Roca & Wang, 1992; Olland & Wang, 1999; Classen *et al.*, 2003). After ATP hydrolysis and release of the resulting ADP and $P_i$, the N-terminal region of the dimer is destabilized (Wigley *et al.*, 1991) and the N gate reopens. The C-terminal region of the ATPase domain has been termed the transducer, which contains important amino acids that intervene in the ATP binding and hydrolysis steps. The transducer was found to signal ATP binding to subsequent domains due to conformational changes following binding of ATP (Roca & Wang, 1992, 1994; Berger *et al.*, 1996; Roca *et al.*, 1996; Harkins & Lindsley, 1998; Harkins *et al.*, 1998; Olland & Wang, 1999; Dong & Berger, 2007). Next to the transducer are the TOPRIM and breakage-reunion domains, which form the DNA gate corresponding to the catalytic site responsible for DNA cleavage. Adjacent to the DNA gate is an α/β fold termed the tower, which also participates in DNA-binding (Dong & Berger, 2007) and from which a long coiled-coil extends, terminating with a globular domain that forms the primary dimer interface, the C gate (Berger *et al.*, 1996; Morais Cabral *et al.*, 1997; Fass *et al.*, 1999).

Type II DNA topoisomerases are thought to operate by a two-gate mechanism (Nitiss, 2009a). According to the proposed model (Figure 6), the topoisomerase binds, through the DNA gate (within the breakage-reunion domain), a segment of double-stranded DNA, the G segment, which will be then cleaved. This is followed by an association with a second segment of DNA, the T segment, and subsequent binding of ATP induces the dimerization of the ATP-binding domains, closing the N gate around the T segment. Hydrolysis of one ATP molecule and the release of its by-products trigger conformational changes and opening of the DNA gate. The T segment is driven through the opening and escapes through an opening of the C gate, while the G segment is resealed. Hydrolysis of the remaining ATP molecule leads to opening of the N gate and resets the enzyme for a new reaction cycle. Once again, ATP hydrolysis is not required for the reversibility of the DNA-topoisomerase covalent bond, but instead seems to be relevant in promoting dimerization and conformational changes, allowing for timed sequential control of topoisomerase action.

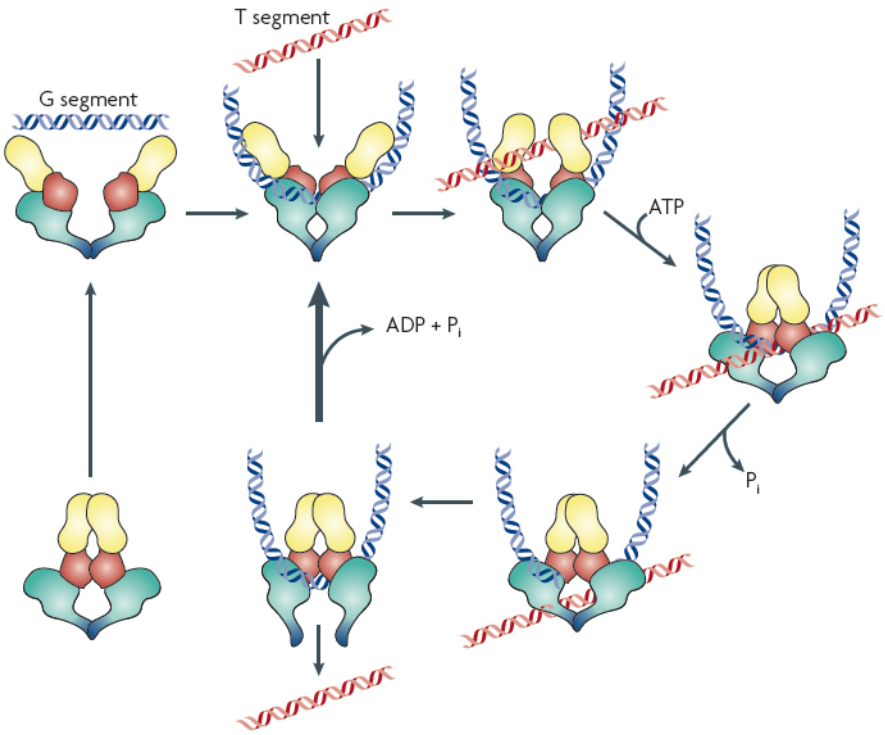


**Figure 6.** Mechanism of strand passage by type II DNA topoisomerases.
Topoisomerase II interacts with two DNA strands to effect strand passage. The enzyme introduces a double-strand break in one DNA strand, termed the G, or gate, strand, and will pass a second strand termed the T, or transported, strand through the break. In the presence of $Mg^{2+}$, the enzyme can cleave the DNA, forming a phosphotyrosine bond between each single strand and a tyrosine in each subunit. ATP binding to the N gate causes the enzyme to form a closed clamp, capturing the T strand that will pass through the break made in the G strand. Afterwards, the T strand exits the enzyme through the C gate. ATP hydrolysis occurs at two steps in the reaction cycle: the first ATP hydrolysed may assist in strand passage, while the second hydrolysis step (along with release of ADP and inorganic phosphate ($P_i$)) allows the clamp to reopen, and allows release of the G segment (for a distributive reaction). Alternately, the enzyme may initiate another catalytic cycle without dissociating from the G strand. Figure and legend adapted from Nitiss (2009a).

### 1.2.4 The biological roles of type II DNA topoisomerases

The central role of type II DNA topoisomerases in solving topological problems arising from DNA replication makes them essential enzymes in all cellular organisms. In both bacteria and eukaryotic cells they intervene in DNA relaxation and decatenation, and in prokaryotic organisms they also participate in DNA supercoiling. They are required as well in other cellular processes, such as transcription and chromosome condensation and separation (Wang, 2002; Nitiss, 2009a; Baxter, 2015).

The vast majority of organisms, from bacteria to mammals, have their DNA globally underwound, or in other words, negatively supercoiled (Schvartzman & Stasiak, 2004). This type of supercoiling facilitates the separation of complementary DNA strands, smoothing the assembly of the machinery required for initiation of DNA replication or transcription (Wang, 2002). The reason for this is that the energetic cost of separating DNA strands is lower when the molecule is negatively supercoiled (Witz & Stasiak, 2010) and, therefore, it comes as no surprise that circular DNA isolated from most organisms is negatively supercoiled (Cozzarelli, 1980). During replication, separation of the two DNA strands by helicases induces superhelical stress on the double helix and accumulation of positively supercoiled DNA ahead of the replication fork, while replicated strands can become interwound behind it (Liu & Wang, 1987; Bermejo *et al.*, 2007; Baxter, 2015) (Figure 7). If left unresolved, these interwindings, also known as precatenanes, can become linked or catenated. While the overwinding (positive supercoiling) ahead of the replication fork can be relaxed by action of a type I or a type II topoisomerase (Wang, 2002), the precatenanes require a topoII for proper

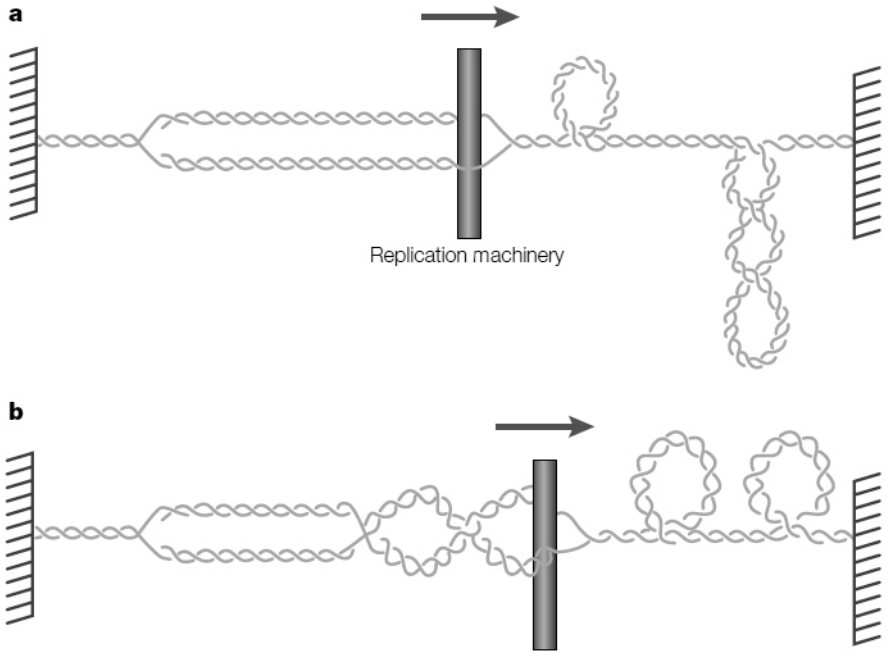


**Figure 7.** Topological problems associated with the progression of the replication machinery.
The replication machinery is illustrated as a rod, and the topological consequences of DNA chain elongation depend on whether the replication machinery can readily rotate in the cellular milieu. DNA ends are attached to a hypothetical immobile structure, which could be the nuclear or cell membrane. a) The replication machinery is immobile; the DNA turns as it is passed through the machinery and positive supercoils accumulate ahead of the advancing replication fork. b) The replication machinery is allowed to rotate around the helical axis of the unreplicated DNA. This rotation allows the redistribution of positive supercoils ahead of the advancing fork into the region behind it, and leads to the intertwining of the pair of duplicated double helices (as depicted) and/or positive supercoiling of that region (not shown). Figure and legend adapted from Wang (2002).

resolution (Peter, Ullsperger, Hiasa, Marians, & Cozzarelli, 1998; Lucas, Germe, Chevrier-Miller, & Hyrien, 2001; Postow *et al.*, 2001). Nevertheless, a study on topoIV from *E. coli* suggests that its action is present throughout the entire DNA replication cycle (Wang *et al.*, 2008), a feature that might be shared by its eukaryotic counterparts. Furthermore, near the end of DNA replication, convergence of opposite replication forks limits the formation of positive supercoils and the action of a type I topoisomerase, resulting in the formation of linked catenanes (Sundin & Varshavsky, 1981) which can only be resolved by action of a type II topoisomerase (Postow *et al.*, 2001; Wang, 2002). Transcription elongation also produces accumulation of positive supercoils ahead of the transcriptional machinery and of negative supercoils behind it (Wang, 2002; Schoeffler & Berger, 2008), but since there is no formation of interlinked DNA molecules, supercoiling is usually resolved by action of type I topoisomerases (Garg, DiAngelo, & Jacob, 1987; Wang, 2002). Still, type II topoisomerases have been shown to substitute for absent topoI and resolve topological problems arising during transcription and/or replication (Brill, DiNardo, Voelkel-Meiman, & Sternglanz, 1987; Kim & Wang, 1989), and for *S. cerevisiae* it has been suggested that Top2 is the main modulator of DNA topology (Salceda, Fernández, & Roca, 2006). In mammals, topoIIβ has been shown to participate in transcriptional regulation, not only being required for transcriptional activation (Ju & Rosenfeld, 2006; Ju *et al.*, 2006) but also for repression (McNamara, Wang, Hanna, & Miller, 2008).

Type II topoisomerases have also been implicated in the regulation of sister chromosome cohesion (Chang, Goulding, Earnshaw, & Carmena, 2003; Carpenter & Porter, 2004; Wang *et al.*, 2008), through the timely resolution of topological entanglements, namely precatenanes and catenanes, originated during DNA replication and hypothesized to contribute, in addition to cohesins, for the maintenance of cohesion between duplicated chromosomes. Indeed, these studies have shown aberrant chromosome separation, in *E. coli* and also several eukaryotic organisms, in the absence of activity of a topoII (more precisely, topoIV or topoIIα, where they exist).

A role in chromosome condensation was suggested for eukaryotic type II topoisomerases, due to studies on fission yeast (Uemura *et al.*, 1987), mammalian cells (Durrieu *et al.*, 2000) and *Xenopus* egg extracts (Cuvier & Hirano, 2003). A likely role of topoII in this process is alleviating the superhelical tension impinged on the DNA by binding of condensin (Kimura & Hirano, 1997; Kimura, Rybenkov, Crisona, Hirano, & Cozzarelli, 1999; Kimura, Cuvier, & Hirano, 2001). Bacterial DNA gyrase has long been implicated in chromosome condensation due to its ability to generate negative supercoiling (Cozzarelli, 1980), the predominant form in which bacterial DNA can be found.

### 1.2.5 Inhibition of type II DNA topoisomerases

Bacterial type II topoisomerases, DNA gyrase and topoisomerase IV, are targets of some of the most efficacious and broad-spectrum antibacterial compounds available, the quinolones. In addition, eukaryotic topoisomerase II is the target of some of the most successful anticancer drugs available (Fortune & Osheroff, 2000; Anderson & Osheroff, 2001; Nitiss, 2009b; Pommier *et al.*, 2010) (Figure 8).

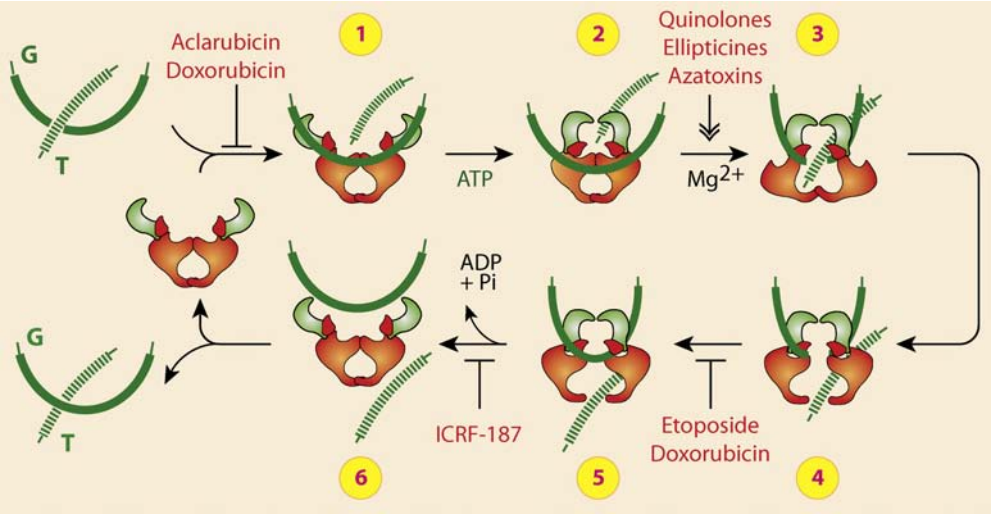


**Figure 8.** Catalytic inhibitors of type II topoisomerases.
Schematic representation of the catalytic cycle of eukaryotic type II topoisomerases. Note that doxorubicin can block the catalytic cycle at two different steps: at low concentration (<1 mM), doxorubicin acts like etoposide (VP-16) by blocking DNA religation (between steps 4 and 5), while at higher concentrations (>10 mM), it interferes with topoII binding to DNA (before step 1). ICRF-187 blocks ATP hydrolysis and inhibits reopening of the ATPase domain (between steps 5 and 6), thereby trapping topological complexes with DNA inside the enzyme. Figure and legend adapted from Pommier *et al.* (2010).

Many drugs that target type II DNA topoisomerases take advantage of the covalent bond between protein and DNA, by inhibiting the re-ligation step and trapping the enzyme in a covalent complex with broken DNA strands (Fortune & Osheroff, 2000) or by favouring the formation of topoisomerase-DNA covalent complexes and, consequently, DNA cleavage (Robinson *et al.*, 1991). These are called topoisomerase II poisons, because they generate breaks (single or double) in DNA and protein covalently bound to DNA. Examples of topoII poisons are the epipodophyllotoxins etoposide (VP-16) and teniposide (VM-26), the anthracyclines daunorubicin and doxorubicin, the acridine amsacrine (*m*-AMSA), quinolones (the majority of which is mostly antibacterial, but some, like CP115,953, affect the activity of eukaryotic topoII), ellipticines, azatoxins and some isoflavones like genistein. TopoII poisons

are thought to interact by interfacial inhibition (Pommier & Cherfils, 2005), a mechanism by which the drug intercalates the DNA after binding of the topoisomerase to it, forming a ternary complex. As the drug also interacts with the protein through hydrogen bonds, the ternary complex is stabilized by the drug and its interactions with both DNA and topoisomerase. This mechanism was recently confirmed for fluoroquinolones moxifloxacin and levofloxacin and their interaction with bacterial topoisomerase IV (Laponogov *et al.*, 2009). Collision of the replication or transcriptional machinery with the drug-induced DNA-topoisomerase covalent complexes generates permanent strand breaks (Nitiss, 2009b) with deleterious effects for the cell.

Other compounds act by inhibiting the catalytic activity of type II topoisomerases without generating covalent complexes, and these are commonly inhibitors of the ATPase domain of topoII (Roca, Ishida, Berger, Andoh, & Wang, 1994; Andoh & Ishida, 1998). As these compounds induce cell death simply by inhibiting the essential enzymatic activity of topoII, they are called topoisomerase II inhibitors. Examples are the bisdioxopiperazines, which include ICRF-187 and ICRF-193, and aminocoumarins, such as novobiocin and coumermycin A1. Curiously, some topoII poisons may also behave as inhibitors, depending on their concentration. This is the case of doxorubicin which, when present in low concentrations traps DNA-topoII covalent complexes, but at high concentrations it intercalates the DNA and inhibits the formation of the covalent complexes (Pommier *et al.*, 2010).

Some mechanisms of resistance to topoisomerase II poisons and inhibitors are similar between prokaryotic and eukaryotic organisms and may involve increased drug efflux or altered drug uptake (Nitiss & Beck, 1996; Jacoby, 2005), due to changes in expression of membrane channels and in order to reduce the intracellular drug concentration. For bacteria, a frequent form of resistance relies on the reduction of the affinity of the drug for its target through stepwise mutagenesis (Jacoby, 2005; Drlica *et al.*, 2009), by altering key amino acids involved in binding of the drug to the topoisomerase. Such mutations frequently lie in a region of the TOPRIM domain that integrates the catalytic pocket, and since quinolones are the main topoII-inhibiting drugs used for antibacterial purposes, the region has been dubbed the quinolone resistance-determining region (QRDR) (Jacoby, 2005; Aldred, Kerns, & Osheroff, 2014). Plasmid-mediated resistance has also been reported (Martínez-Martínez, Pascual, & Jacoby, 1998), due to expression of the *qnr* gene coding for a 219 amino acid protein, belonging to the pentapeptide repeat family (Shah & Heddle, 2014), that binds bacterial type II topoisomerases and protects them from quinolone action (Tran, Jacoby, & Hooper, 2005). It does not, however, protect against the action of inhibitors of the ATPase domain, like coumermycin A1. In eukaryotic cells, resistance to topoII-inhibiting drugs may

also arise from reduction of protein levels (less topoisomerase leads to a reduction in covalent complexes and, thus, to less DNA damage) (Beck, Danks, Wolverton, Kim, & Chen, 1993), changes in subcellular localization (Wessel *et al.*, 1997; Engel *et al.*, 2004), regulation of enzymatic activity by post-translational modifications such as phosphorylation (Isaacs *et al.*, 1998) or, as in their bacterial counterparts, accumulation of specific mutations in the catalytic pocket (Nitiss & Beck, 1996) that hamper binding of the drug.

### 1.3 Aim of the work

African swine fever virus ORF P1192R was described as putatively coding for a protein with significant sequence homology to type II DNA topoisomerases (Baylis, Dixon, Vydelingum, & Smith, 1992; García-Beato *et al.*, 1992a), based on genomic and bioinformatics analyses. Despite the time passed since its description, its confirmation and characterization are still lacking. Therefore, the aim of this work was to confirm pP1192R as a viral type II DNA topoisomerase, to characterize the biochemical properties and enzymatic activity and to study its expression throughout the viral replication cycle and its cellular localization and function.

Specifically, the objectives were:

- Perform comprehensive bioinformatics and phylogenetic analyses of P1192R;
- To study the biologic activity of recombinant pP1192R, through cloning, expression, purification and characterization of its catalytic activity *in vitro*;
- To screen the effect of topoisomerase II poisons and inhibitors on the *in vitro* activity of recombinant pP1192R;
- To assess protein levels and intracellular localization of pP1192R in ASFV-infected cells, at different time-points of infection;
- To evaluate the effect of mutations, generated by site-directed mutagenesis, in regulatory residues or domains of pP1192R predicted by the bioinformatics analyses mentioned above.

## 1.4 References


Alcamí, a, Angulo, a, López-Otín, C., Muñoz, M., Freije, J. M., Carrascosa, a L., & Viñuela, E. (1992). Amino acid sequence and structural properties of protein p12, an African swine fever virus attachment protein. *Journal of Virology* , *66* (6 ), 3860–3868.

Aldred, K. J., Kerns, R. J., & Osheroff, N. (2014). Mechanism of Quinolone Action and Resistance. *Biochemistry*, *53*(10), 1565–1574.

Anderson, E., Hutchings, G. H., Mukarati, N., & Wilkinson, P. J. (1998). African swine fever virus infection of the bushpig (Potamochoerus porcus) and its significance in the epidemiology of the disease. *Veterinary Microbiology*, *62*(1), 1–15.

Anderson, V. E., & Osheroff, N. (2001). Type II topoisomerases as targets for quinolone antibacterials: turning Dr. Jekyll into Mr. Hyde. *Current Pharmaceutical Design*, *7*(5), 337–53.

Andoh, T., & Ishida, R. (1998). Catalytic inhibitors of DNA topoisomerase II. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1400*(1-3), 155–171.

Andrés, G., Alejo, A., Salas, J., & Salas, M. L. (2002). African Swine Fever Virus Polyproteins pp220 and pp62 Assemble into the Core Shell. *Journal of Virology*, *76*(24), 12473–12482.

Andrés, G., Alejo, A., Simón-Mateo, C., & Salas, M. L. (2001a). African swine fever virus protease, a new viral member of the SUMO-1-specific protease family. *Journal of Biological Chemistry*, *276*(1), 780–787.

Andrés, G., García-Escudero, R., Simón-Mateo, C., & Viñuela, E. (1998). African swine fever virus is enveloped by a two-membraned collapsed cisterna derived from the endoplasmic reticulum. *Journal of Virology*, *72*(11), 8988–9001.

Andrés, G., García-Escudero, R., Viñuela, E., Salas, M. L., Rodríguez, J. M., García-Escudero, R., Viñuela, E., Salas, M. L., & Rodríguez, J. M. (2001b). African swine fever virus structural protein pE120R is essential for virus transport from assembly sites to plasma membrane but not for infectivity. *Journal of Virology*, *75*(15), 6758–6768.

Andrés, G., Simón-Mateo, C., & Viñuela, E. (1997). Assembly of African swine fever virus: role of polyprotein pp220. *Journal of Virology* , *71* (3 ), 2331–2341.

Aravind, L., Leipe, D. D., & Koonin, E. V. (1998). Toprim - A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*, *26*(18), 4205–4213.

Arias, M., & Sánchez-Vizcaíno, J. M. (2008). African Swine Fever. In *Trends in Emerging Viral Infections of Swine* (pp. 119–124). Ames, Iowa, USA: Iowa State Press.

Ballester, M., Galindo-Cardiel, I., Gallardo, C., Argilaguet, J. M., Segalés, J., Rodríguez, J. M., & Rodríguez, F. (2010). Intranuclear detection of African swine fever virus DNA in several cell types from formalin-fixed and paraffin-embedded tissues using a new in situ hybridisation protocol. *Journal of Virological Methods*, *168*(1-2), 38–43.

Baroudy, B. M., Venkatesan, S., & Moss, B. (1982). Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell*, *28*(2), 315–324.

Baroudy, B. M., Venkatesan, S., & Moss, B. (1983). Structure and Replication of Vaccinia Virus Telomeres. *Cold Spring Harbor Symposia on Quantitative Biology*, *47*(0), 723–729.

Basto, A. P., Nix, R. J., Boinas, F., Mendes, S., Silva, M. J., Cartaxeiro, C., Portugal, R. S., Leitão, A., Dixon, L. K., & Martins, C. (2006). Kinetics of African swine fever virus infection in Ornithodoros erraticus ticks. *The Journal of General Virology*, *87*(Pt 7), 1863–71.

Bastos, A. D. S., Penrith, M.-L. M.-L., Crucière, C., Edrich, J. L., Hutchings, G., Roger, F., Couacy-Hymann, E., R Thomson, G., & R.Thomson, G. (2003). Genotyping field strains of African swine fever virus by partial p72 gene characterisation. *Archives of Virology*, *148*(4), 693–706.

Baxter, J. (2015). "Breaking Up Is Hard to Do": The Formation and Resolution of Sister Chromatid Intertwines. *Journal of Molecular Biology*, *427*(3), 590–607.

Baylis, S. A., Dixon, L. K., Vydelingum, S., & Smith, G. L. (1992). African swine fever virus encodes a gene with extensive homology to type II DNA topoisomerases. *Journal of Molecular Biology*, *228*(3), 1003–1010.

Beck, W. T., Danks, M. K., Wolverton, J. S., Kim, R., & Chen, M. (1993). Drug resistance associated with altered DNA topoisomerase II. *Advances in Enzyme Regulation*, *33*(0), 113–116.

Beltrán-Alcrudo, D., Lubroth, J., Depner, K., & De La Rocque, S. (2008). African swine fever in the Caucasus. *FAO Empres Watch*, 1–8.

Berger, J. M., Gamblin, S. J., Harrison, S. C., & Wang, J. C. (1996). Structure and mechanism of DNA topoisomerase II. *Nature*.

Bergerat, A., Gadelle, D., & Forterre, P. (1994). Purification of a DNA topoisomerase II from the hyperthermophilic archaeon Sulfolobus shibatae. A thermostable enzyme with both bacterial and eucaryal features. *Journal of Biological Chemistry* , *269* (44 ), 27663–27669.

Bermejo, R., Doksani, Y., Capra, T., Katou, Y.-M., Tanaka, H., Shirahige, K., & Foiani, M. (2007). Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes & Development*, *21*(15), 1921–1936.

Bishop, R. P., Fleischauer, C., de Villiers, E. P., Okoth, E. a., Arias, M., Gallardo, C., & Upton, C. (2015). Comparative analysis of the complete genome sequences of Kenyan African swine fever virus isolates within p72 genotypes IX and X. *Virus Genes*, *50*(2), 303–309.

Blasco, R., Agüero, M., Almendral, J., & Viñuela, E. (1989). Variable and constant regions in african swine fever virus DNA. *Virology*, *168*(2), 330–338.

Blome, S., Gabriel, C., & Beer, M. (2013). Pathogenesis of African swine fever in domestic pigs and European wild boar. *Virus Research*, *173*(1), 122–30.

Boinas, F. S., Wilson, A. J., Hutchings, G. H., Martins, C., & Dixon, L. J. (2011). The Persistence of African Swine Fever Virus in Field-Infected Ornithodoros erraticus during the ASF Endemic Period in Portugal. *PloS One*, *6*(5), e20383.

Borca, M. V, Irusta, P. M., Kutish, G. F., Carrillo, C., Afonso, C. L., Burrage, T., Neilan, J. G., & Rock, D. L. (1996). A structural DNA binding protein of African swine fever virus with similarity to bacterial histone-like proteins. *Archives of Virology*, *141*(2), 301–313.

Boshoff, C. I., Bastos, A. D. S., Gerber, L. J., & Vosloo, W. (2007). Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973–1999). *Veterinary Microbiology*, *121*(1–2), 45–55.

Breese, S. S. J., & DeBoer, C. J. (1966). Electron microscope observations of African swine fever virus in tissue culture cells. *Virology*, *28*(3), 420–428.

Brill, S. J., DiNardo, S., Voelkel-Meiman, K., & Sternglanz, R. (1987). Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature*, *326*(6111), 414–416.

Brookes, S. M., Hyatt, A. D., Wise, T., & Parkhouse, R. M. E. (1998). Intracellular virus DNA distribution and the acquisition of the nucleoprotein core during African swine fever virus particle assembly: ultrastructural in situ hybridisation and DNase-gold labelling. *Virology*, *249*(1), 175–188.

Brown, P. O., & Cozzarelli, N. R. (1981). Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(2), 843–847.

Burrage, T. G. (2013). African swine fever virus infection in Ornithodoros ticks. *Virus Research*, *173*(1), 131–139.

Caeiro, F., Meireles, M., Ribeiro, G., & Costa, J. V. (1990). In vitro DNA replication by cytoplasmic extracts from cells infected with african swine fever virus. *Virology*, *179*(1), 87–94.

Caiado, J. M., Boinas, F. S., & Louza, A. C. (1988). Epidemiological research of African swine fever (ASF) in Portugal: the role of vectors and virus reservoirs. *Acta Veterinaria Scandinavia. Supplementum*, *84*, 136–138.

Carpenter, A. J., & Porter, A. C. G. (2004). Construction, Characterization, and Complementation of a Conditional-Lethal DNA Topoisomerase IIα Mutant Human Cell Line. *Molecular Biology of the Cell* , *15* (12 ), 5700–5711.

Carrascosa, A. L., Saastre, I., Gonzalez, P., & Viñuela, E. (1993). Localization of the African swine fever virus attachment protein P12 in the virus particle by immunoelectron microscopy. *Virology*, *193*(1), 460–465.

Carrascosa, J. L., Carazo, J. M., Carrascosa, A. L., García, N., Santisteban, A., & Viñuela, E. (1984). General morphology and capsid fine structure of African swine fever virus particles. *Virology*, *132*(1), 160–172.

Champoux, J. J. (1981). DNA is linked to the rat liver DNA nicking-closing enzyme by a phosphodiester bond to tyrosine. *Journal of Biological Chemistry* , *256* (10 ), 4805–4809.

Chang, C.-J., Goulding, S., Earnshaw, W. C., & Carmena, M. (2003). RNAi analysis reveals an unexpected role for topoisomerase II in chromosome arm congression to a metaphase plate. *Journal of Cell Science* , *116* (23 ), 4715–4726.

Chapman, D. a G., Tcherepanov, V., Upton, C., & Dixon, L. K. (2008). Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *The Journal of General Virology*, *89*(Pt 2), 397–408.

Cheng, C., Kussie, P., Pavletich, N., & Shuman, S. (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, *92*(6), 841–850.

Classen, S., Olland, S., & Berger, J. M. (2003). Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 10629–34.

Cobbold, C., Windsor, M., & Wileman, T. (2001). A Virally Encoded Chaperone Specialized for Folding of the Major Capsid Protein of African Swine Fever Virus. *Journal of Virology* , *75* (16 ), 7221–7229.

Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D. K., Cheng, X.-W., Federici, B. A., Van Etten, J. L., Koonin, E. V, La Scola, B., & Raoult, D. (2013). "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Archives of Virology*, *158*(12), 2517–21.

Costard, S., Mur, L., Lubroth, J., Sanchez-Vizcaino, J. M., & Pfeiffer, D. U. (2013). Epidemiology of African swine fever virus. *Virus Research*, *173*(1), 191–197.

Cozzarelli, N. (1980). DNA gyrase and the supercoiling of DNA. *Science*, *207*(4434), 953–960.

Cuvier, O., & Hirano, T. (2003). A role of topoisomerase II in linking DNA replication to chromosome condensation. *The Journal of Cell Biology*, *160*(5), 645–655.

De Villiers, E. P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R., & Bishop, R. P. (2010). Phylogenomic analysis of 11 complete African swine fever virus genome sequences. *Virology*, *400*(1), 128–36.

Dixon, L. K., Alonso, C., Escribano, J. M., Martins, C., Revilla, Y., Salas, M. L., & Takamatsu, H. (2012). Family Asfarviridae. In C. R. P. & R. B. W. M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo (Ed.), *Virus Taxonomy: Classification*

*and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses* (pp. 153–162). San Diego:Academic Press.

Dixon, L. K., Chapman, D. A. G., Netherton, C. L., & Upton, C. (2013). African swine fever virus replication and genomics. *Virus Research*, *173*(1), 3–14.

Dixon, L. K., & Wilkinson, P. J. (1988). Genetic Diversity of African Swine Fever Virus Isolates from Soft Ticks (Ornithodoros moubata) Inhabiting Warthog Burrows in Zambia. *Journal of General Virology*, *69*(12), 2981–2993.

Domanico, P. L., & Tse-Dinh, Y. C. (1991). Mechanistic studies on E. coli DNA topoisomerase I: Divalent ion effects. *Journal of Inorganic Biochemistry*, *42*(2), 87–96.

Dong, K. C., & Berger, J. M. (2007). Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature*, *450*(7173), 1201–1205.

Drlica, K. (1992). Control of bacterial DNA supercoiling. *Molecular Microbiology*, *6*(4), 425–433.

Drlica, K., Hiasa, H., Kerns, R., Malik, M., Mustaev, A., & Zhao, X. (2009). Quinolones: Action and Resistance Updated . *Current Topics in Medicinal Chemistry*, *9*(11), 981–998.

Durrieu, F., Samejima, K., Fortune, J. M., Kandels-Lewis, S., Osheroff, N., & Earnshaw, W. C. (2000). DNA topoisomerase IIα interacts with CAD nuclease and is involved in chromatin condensation during apoptotic execution. *Current Biology*, *10*(15), 923–S2.

Dynan, W. S., Jendrisak, J. J., Hager, D. A., & Burgess, R. R. (1981). Purification and characterization of wheat germ DNA topoisomerase I (nicking-closing enzyme). *Journal of Biological Chemistry* , *256* (11 ), 5860–5865.

EFSA AHAW Panel. (2014). Scientific Opinion on African swine fever. *EFSA Journal*, *12*(4), 3628.

Engel, R., Valkov, N. I., Gump, J. L., Hazlehurst, L., Dalton, W. S., & Sullivan, D. M. (2004). The cytoplasmic trafficking of DNA topoisomerase IIα correlates with etoposide resistance in human myeloma cells. *Experimental Cell Research*, *295*(2), 421–431.

Epifano, C., Krijnse-Locker, J., Salas, M. L., Rodríguez, J. M., & Salas, J. (2006a). The African Swine Fever Virus Nonstructural Protein pB602L Is Required for Formation of the Icosahedral Capsid of the Virus Particle. *Journal of Virology* , *80* (24 ), 12260–12270.

Epifano, C., Krijnse-Locker, J., Salas, M. L., Salas, J., & Rodríguez, J. M. (2006b). Generation of Filamentous Instead of Icosahedral Particles by Repression of African Swine Fever Virus Structural Protein pB438L. *Journal of Virology* , *80* (23 ), 11456–11466.

Esteves, A., Marques, M. I., & Costa, J. V. (1986). Two-dimensional analysis of African swine fever virus proteins and proteins induced in infected cells. *Virology*, *152*(1), 192–206.

Estrada-Pena, A., Farkas, R., Jaenson, T. G. T., Madder, M., Pascucci, I., & Tarrés-Call, J. (2010). Scientific opinion on the Role of Tick Vectors in the Epidemiology of Crimean-Congo Hemorrhagic Fever and African Swine Fever in Eurasia: EFSA Panel on Animal Health and Welfare.

Fass, D., Bogden, C. E., & Berger, J. M. (1999). Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands. *Nature Structural Biology*, *6*(4), 322–326.

Forterre, P., Bergerat, A., & Lopex-Garcia, P. (1996). The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiology Reviews*, *18*(2-3), 237–248.

Forterre, P., Gribaldo, S., Gadelle, D., & Serre, M.-C. (2007). Origin and evolution of DNA topoisomerases. *Biochimie*, *89*(4), 427–46.

Fortune, J. M., & Osheroff, N. (2000). Topoisomerase II as a target for anticancer drugs: when enzymes stop being nice. *Progress in Nucleic Acid Research and Molecular Biology*, *64*, 221–253.

Fritsche, M., Li, S., Heermann, D. W., & Wiggins, P. A. (2012). A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Research*, *40*(3), 972–980.

Gadelle, D., Filée, J., Buhler, C., & Forterre, P. (2003). Phylogenomics of type II DNA topoisomerases. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *25*(3), 232–42.

Gago da Câmara, N. J. (1932). História da peste suína em Angola. *Pecuária*, *1*.

Gallardo, C., Mwaengo, D., Macharia, J., Arias, M., Taracha, E., Soler, A., Okoth, E., Martín, E., Kasiti, J., & Bishop, R. (2009). Enhanced discrimination of African swine fever virus isolates through nucleotide sequencing of the p54, p72, and pB602L (CVR) genes. *Virus Genes*, *38*(1), 85–95.

García-Beato, R., Freije, J. M., López-Otín, C., Blasco, R., Viñuela, E., & Salas, M. L. (1992a). A gene homologous to topoisomerase II in African swine fever virus. *Virology*, *188*(2), 938–47.

García-Beato, R., Salas, M. L., Viñuela, E., & Salas, J. (1992b). Role of the host cell nucleus in the replication of African swine fever virus DNA. *Virology*, *188*(2), 637–649.

García-Escudero, R., Andrés, G., Almazán, F., & Viñuela, E. (1998). Inducible Gene Expression from African Swine Fever Virus Recombinants: Analysis of the Major Capsid Protein p72. *Journal of Virology* , *72* (4 ), 3185–3195.

Garg, L. C., DiAngelo, S., & Jacob, S. T. (1987). Role of DNA topoisomerase I in the transcription of supercoiled rRNA gene. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(10), 3185–3188.

Gellert, M., Mizuuchi, K., O'Dea, M. H., & Nash, H. A. (1976a). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(11), 3872–3876.

Gellert, M., O'Dea, M. H., Itoh, T., & Tomizawa, J. (1976b). Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(12), 4474–4478.

González, A., Talavera, A., Almendral, J. M., & Viñuela, E. (1986). Hairpin loop structure of African swine fever virus DNA. *Nucleic Acids Research*, *14*(17), 6835–44.

Goto, T., & Wang, J. C. (1984). Yeast DNA topoisomerase II is encoded by a single-copy, essential gene. *Cell*, *36*(4), 1073–80.

Grue, P., Grässer, a, Sehested, M., Jensen, P. B., Uhse, a, Straub, T., Ness, W., & Boege, F. (1998). Essential mitotic functions of DNA topoisomerase IIalpha are not adopted by topoisomerase IIbeta in human H69 cells. *The Journal of Biological Chemistry*, *273*(50), 33660–6.

Haresnape, J. M., & Mamu, F. D. (1986). The distribution of ticks of the *Ornithodoros moubata* complex (Ixodoidea: Argasidae) in Malawi, and its relation to African swine fever epizootiology. *Epidemiology & Infection*, *96*(03), 535–544.

Haresnape, J. M., & Wilkinson, P. J. (1989). A study of African swine fever virus infected ticks ( *Ornithodoros moubata*) collected from three villages in the ASF enzootic area of Malawi following an outbreak of the disease in domestic pigs. *Epidemiology & Infection*, *102*(03), 507–522.

Haresnape, J. M., Wilkinson, P. J., & Mellor, P. S. (1988). Isolation of African swine fever virus from ticks of the *Ornithodoros moubata* complex (Ixodoidea: Argasidae) collected within the African swine fever enzootic area of Malawi. *Epidemiology & Infection*, *101*(01), 173–185.

Harkins, T. T., Lewis, T. J., & Lindsley, J. E. (1998). Pre-Steady-State Analysis of ATP Hydrolysis by Saccharomyces cerevisiae DNA Topoisomerase II. 2. Kinetic Mechanism for the Sequential Hydrolysis of Two ATP †. *Biochemistry*, *37*(20), 7299–7312.

Harkins, T. T., & Lindsley, J. E. (1998). Pre-Steady-State Analysis of ATP Hydrolysis by Saccharomyces cerevisiae DNA Topoisomerase II. 1. A DNA-Dependent Burst in ATP Hydrolysis. *Biochemistry*, *37*(20), 7292–7298.

Isaacs, R. J., Davies, S. L., Sandri, M. I., Redwood, C., Wells, N. J., & Hickson, I. D. (1998). Physiological regulation of eukaryotic topoisomerase II. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1400*(1-3), 121–137.

Iyer, L. M., Aravind, L., & Koonin, E. V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. *Journal of Virology*, *75*(23), 11720–11734.

Jacoby, G. A. (2005). Mechanisms of Resistance to Quinolones. *Clinical Infectious Diseases*, *41*(Supplement 2), S120–S126.

Jori, F., & Bastos, A. (2009). Role of Wild Suids in the Epidemiology of African Swine Fever. *EcoHealth*, *6*(2), 296–310.

Jori, F., Vial, L., Penrith, M. L., Pérez-Sánchez, R., Etter, E., Albina, E., Michaud, V., & Roger, F. (2013). Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. *Virus Research*, *173*(1), 212–227.

Jouvenet, N., Monaghan, P., Way, M., & Wileman, T. (2004). Transport of African Swine Fever Virus from Assembly Sites to the Plasma Membrane Is Dependent on Microtubules and Conventional Kinesin. *Journal of Virology* , *78* (15 ), 7990–8001.

Ju, B.-G., Lunyak, V. V, Perissi, V., Garcia-Bassets, I., Rose, D. W., Glass, C. K., & Rosenfeld, M. G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science*, *312*(5781), 1798–1802.

Ju, B.-G., & Rosenfeld, M. G. (2006). A Breaking Strategy for Topoisomerase IIβ/PARP-1-Dependent Regulated Transcription. *Cell Cycle*, *5*(22), 2557–2560.

Kanaar, R., & Cozzarelli, N. R. (1992). Roles of supercoiled DNA structure in DNA transactions. *Current Opinion in Structural Biology*, *2*(3), 369–379.

Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., & Suzuki, H. (1990). New topoisomerase essential for chromosome segregation in E. coli. *Cell*, *63*(2), 393–404.

Kim, R., & Wang, J. C. (1989). Function of DNA topoisomerases as replication swivels in Saccharomyces cerevisiae. *Journal of Molecular Biology*, *208*(2), 257–267.

Kimura, K., Cuvier, O., & Hirano, T. (2001). Chromosome Condensation by a Human Condensin Complex in Xenopus Egg Extracts. *Journal of Biological Chemistry*, *276*(8), 5417–5420.

Kimura, K., & Hirano, T. (1997). ATP-Dependent Positive Supercoiling of DNA by 13S Condensin: A Biochemical Implication for Chromosome Condensation. *Cell*, *90*(4), 625–634.

Kimura, K., Rybenkov, V. V, Crisona, N. J., Hirano, T., & Cozzarelli, N. R. (1999). 13S Condensin Actively Reconfigures DNA by Introducing Global Positive Writhe: Implications for Chromosome Condensation. *Cell*, *98*(2), 239–248.

Kleiboeker, S. B., & Scoles, G. A. (2001). Pathogenesis of African swine fever virus in *Ornithodoros* ticks. *Animal Health Research Reviews*, *2*(02), 121–128.

Kramlinger, V. M., & Hiasa, H. (2006). The “GyrA-box” Is Required for the Ability of DNA Gyrase to Wrap DNA and Catalyze the Supercoiling Reaction. *Journal of Biological Chemistry* , *281* (6 ), 3738–3742.

Kreuzer, K. N. (1998). Bacteriophage T4, a model system for understanding the mechanism of type II topoisomerase inhibitors. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1400*(1-3), 339–347.

Laponogov, I., Sohi, M. K., Veselkov, D. A., Pan, X.-S., Sawhney, R., Thompson, A. W., McAuley, K. E., Fisher, L. M., & Sanderson, M. R. (2009). Structural insight into the

quinolone–DNA cleavage complex of type IIA topoisomerases. *Nature Structural & Molecular Biology*, *16*(6), 667–669.

Lindsley, J. E., & Wang, J. C. (1991). Proteolysis patterns of epitopically labeled yeast DNA topoisomerase II suggest an allosteric transition in the enzyme induced by ATP binding. *Proceedings of the National Academy of Sciences*, *88*(23), 10485–10489.

Linka, R. M., Porter, A. C. G., Volkov, A., Mielke, C., Boege, F., & Christensen, M. O. (2007). C-terminal regions of topoisomerase IIalpha and IIbeta determine isoform-specific functioning of the enzymes in vivo. *Nucleic Acids Research*, *35*(11), 3810–22.

Liu, L. F., Liu, C., & Alberts, B. M. (1980). Type II DNA Topoisomerases : Enzymes That Can Unknot a Topologically Knotted DNA Molecule via a Reversible Double-Strand Break, *19*(March), 697–707.

Liu, L. F., Rowe, T. C., Yang, L., Tewey, K. M., & Chen, G. L. (1983). Cleavage of DNA by mammalian DNA topoisomerase II. *Journal of Biological Chemistry*, *258*(24), 15365–15370.

Liu, L. F., & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(20), 7024–7027.

Lubisi, B. A., Bastos, A. D. S., Dwarka, R. M., & Vosloo, W. (2005). Molecular epidemiology of African swine fever in East Africa. *Archives of Virology*, *150*(12), 2439–2452.

Lucas, I., Germe, T., Chevrier-Miller, M., & Hyrien, O. (2001). Topoisomerase II can unlink replicating DNA by precatenane removal. *EMBO Journal*, *20*(22), 6509–6519.

Manso Ribeiro, J. J., & Azevedo, J. A. (1961). La peste porcine africaine au Portugal. *Bulletin de l'Office Internationale Des Épizooties*, (55), 88–108.

Martínez-Martínez, L., Pascual, A., & Jacoby, G. A. (1998). Quinolone resistance from a transferable plasmid. *The Lancet*, *351*(9105), 797–799.

Matos, A. P. A., & Carvalho, Z. G. (1993). African swine fever virus interaction with microtubules. *Biology of the Cell*, *78*(3), 229–234.

McClendon, A. K., Rodriguez, A. C., & Osheroff, N. (2005). Human topoisomerase IIα rapidly relaxes positively supercoiled DNA: Implications for enzyme action ahead of replication forks. *Journal of Biological Chemistry*, *280*(47), 39337–39345.

McNamara, S., Wang, H., Hanna, N., & Miller, W. H. (2008). Topoisomerase II Negatively Modulates Retinoic Acid Receptor Function: a Novel Mechanism of Retinoic Acid Resistance. *Molecular and Cellular Biology*, *28*(6), 2066–2077.

Mirski, S. E., Gerlach, J. H., Cummings, H. J., Zirngibl, R., Greer, P. a, & Cole, S. P. (1997). Bipartite nuclear localization signals in the C terminus of human topoisomerase II alpha. *Experimental Cell Research*, *237*(2), 452–5.

Mirski, S. E. L., Gerlach, J. H., & Cole, S. P. C. (1999). Sequence Determinants of Nuclear Localization in the α and β Isoforms of Human Topoisomerase II. *Experimental Cell Research*, *339*, 329–339.

Montgomery, R. E. (1921). On A Form of Swine Fever Occurring in British East Africa (Kenya Colony). *Journal of Comparative Pathology and Therapeutics*, *34*(0), 159–191.

Morais Cabral, J. H., Jackson, a P., Smith, C. V, Shikotra, N., Maxwell, a, & Liddington, R. C. (1997). Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature*, *388*(6645), 903–906.

Morrison, A., & Cozzarelli, N. R. (1979). Site-specific cleavage of DNA by E. coli DNA gyrase. *Cell*, *17*(1), 175–184.

Moss, B. (2013). Poxvirus DNA Replication. *Cold Spring Harbor Perspectives in Biology*, *5*(9), a010199–a010199.

Muñoz, M., Freije, J. M. P., Salas, M. L., Viñuela, E., & López-Otín, C. (1993). Structure and expression in E. coli of the gene coding for protein p10 of African swine fever virus. *Archives of Virology*, *130*(1-2), 93–107.

Nitiss, J. L. (2009a). DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews. Cancer*, *9*(5), 327–37.

Nitiss, J. L. (2009b). Targeting DNA topoisomerase II in cancer chemotherapy. *Nature Reviews. Cancer*, *9*(5), 338–350.

Nitiss, J. L., & Beck, W. T. (1996). Antitopoisomerase drug action and resistance. *European Journal of Cancer*, *32*(6), 958–966.

Nolan, J. M., Lee, M. P., Wyckoff, E., & Hsieh, T. S. (1986). Isolation and characterization of the gene encoding Drosophila DNA topoisomerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(11), 3664–3668.

Oleaga-Pérez, A., Pérez-Sánchez, R., & Encinas-Grandes, A. (1990). Distribution and biology of Ornithodoros erraticus in parts of Spain affected by African swine fever. *The Veterinary Record*, *126*(2), 32–37.

Olland, S., & Wang, J. C. (1999). Catalysis of ATP Hydrolysis by Two NH2-terminal Fragments of Yeast DNA Topoisomerase II. *Journal of Biological Chemistry*, *274*(31), 21688–21694.

Ortin, J., & Viñuela, E. (1977). Requirement of cell nucleus for African swine fever virus replication in Vero cells. *Journal of Virology*, *21*(3), 902–905.

Osheroff, N. (1987). Role of the divalent cation in topoisomerase II mediated reactions. *Biochemistry*, *26*(20), 6402–6406.

Osheroff, N., Shelton, E. R., & Brutlag, D. L. (1983). DNA topoisomerase II from Drosophila melanogaster. Relaxation of supercoiled DNA. *The Journal of Biological Chemistry*, *258*(15), 9536–9543.

Osheroff, N., & Zechiedrich, E. L. (1987). Calcium-promoted DNA cleavage by eukaryotic topoisomerase II: trapping the covalent enzyme-DNA complex in an active form. *Biochemistry*, *26*(14), 4303–4309.

Peng, H., & Marians, K. J. (1993). Escherichia coli topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. *Journal of Biological Chemistry*, *268*(32), 24481–24490.

Pérez-Sánchez, R., Astigarraga, A., Oleaga-Perez, A., & Encinas-Grandes, A. (1994). Relationship between the persistence of African swine fever and the distribution of Ornithodoros erraticus in the province of Salamanca, Spain. *The Veterinary Record*, *135*(9), 207–209.

Peter, B. J., Ullsperger, C., Hiasa, H., Marians, K. J., & Cozzarelli, N. R. (1998). The Structure of Supercoiled Intermediates in DNA Replication. *Cell*, *94*(6), 819–827.

Plowright, W. (1986). African swine fever: a retrospective view. *Revue Scientifique et Technique de l'OIE*, *5*(2), 455–468.

Plowright, W., Perry, C. T., & Greig, A. (1974). Sexual transmission of African swine fever virus in the tick, Ornithodoros moubata porcinus, Walton. *Research in Veterinary Science*, *17*(1), 106.

Plowright, W., Perry, C. T., & Peirce, M. A. (1970). Transovarial infection with African swine fever virus in the argasid tick, Ornithodoros moubata porcinus, Walton. *Research in Veterinary Science*, *11*, 582–584.

Pommier, Y., & Cherfils, J. (2005). Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends in Pharmacological Sciences*, *26*(3), 138–145.

Pommier, Y., Leo, E., Zhang, H., & Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chemistry & Biology*, *17*(5), 421–33.

Portugal, R., Coelho, J., Hoper, D., Little, N. S., Smithson, C., Upton, C., Martins, C., Leitao, A., & Keil, G. M. (2015). Related strains of African swine fever virus with different virulence: genome comparison and analysis. *Journal of General Virology*, *96*(Pt_2), 408–419.

Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., & Cozzarelli, N. R. (2001). Topological challenges to DNA replication: Conformations at the fork. *Proceedings of the National Academy of Sciences*, *98*(15), 8219–8226.

Postow, L., Hardy, C. D., Arsuaga, J., & Cozzarelli, N. R. (2004). Topological domain structure of the Escherichia coli chromosome. *Genes & Development*, *18*(14), 1766–1779.

Rennie, L., Wilkinson, P. J., & Mellor, P. S. (2001). Transovarial transmission of African swine fever virus in the argasid tick Ornithodoros moubata. *Medical and Veterinary Entomology*, *15*(2), 140–146.

Robinson, M. J., Martin, B. A., Gootz, T. D., McGuirk, P. R., Moynihan, M., Sutcliffe, J. A., & Osheroff, N. (1991). Effects of quinolone derivatives on eukaryotic topoisomerase II. A novel mechanism for enhancement of enzyme-mediated DNA cleavage. *Journal of Biological Chemistry*, *266*(22), 14585–14592.

Roca, J., Berger, J. M., Harrison, S. C., & Wang, J. C. (1996). DNA transport by a type II topoisomerase: direct evidence for a two-gate mechanism. *Proceedings of the National Academy of Sciences*, *93*(9), 4057–4062.

Roca, J., Ishida, R., Berger, J. M., Andoh, T., & Wang, J. C. (1994). Antitumor bisdioxopiperazines inhibit yeast DNA topoisomerase II by trapping the enzyme in the form of a closed protein clamp. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(5), 1781–1785.

Roca, J., & Wang, J. C. (1992). The capture of a DNA double helix by an ATP-dependent protein clamp: A key step in DNA transport by type II DNA topoisomerases. *Cell*, *71*(5), 833–840.

Roca, J., & Wang, J. C. (1994). DNA transport by a type II DNA topoisomerase: Evidence in favor of a two-gate mechanism. *Cell*, *77*(4), 609–616.

Roca, J., & Wang, J. C. (1996). The probabilities of supercoil removal and decatenation by yeast DNA topoisomerase II. *Genes to Cells*, *1*(1), 17–27.

Rodríguez, I., Nogal, M. L., Redrejo-Rodríguez, M., Bustos, M. J., & Salas, M. L. (2009). The African Swine Fever Virus Virion Membrane Protein pE248R Is Required for Virus Infectivity and an Early Postentry Event. *Journal of Virology* , *83* (23 ), 12290–12300.

Rodríguez, J. M., García-Escudero, R., Salas, M. L., Andrés, G., Rodriguez, J. M., Garcia-Escudero, R., Salas, M. L., Andres, G., Rodríguez, J. M., García-Escudero, R., Salas, M. L., Andrés, G., Rodriguez, J. M., Garcia-Escudero, R., Salas, M. L., & Andres, G. (2004). African Swine Fever Virus Structural Protein p54 Is Essential for the Recruitment of Envelope Precursors to Assembly Sites. *Journal of Virology*, *78*(8), 4299–4313.

Rodríguez, J. M., & Salas, M. L. (2013). African swine fever virus transcription. *Virus Research*.

Rodríguez, J. M., Yáñez, R. J., Almazán, F., Viñuela, E., & Rodriguez, J. F. (1993). African swine fever virus encodes a CD2 homolog responsible for the adhesion of erythrocytes to infected cells. *Journal of Virology* , *67* (9 ), 5312–5320.

Rojo, G., García-Beato, R., Viñuela, E., Salas, M. L., & Salas, J. (1999). Replication of African swine fever virus DNA in infected cells. *Virology*, *257*(2), 524–36.

Rouiller, I., Brookes, S. M., Hyatt, A. D., Windsor, M., & Wileman, T. (1998). African Swine Fever Virus Is Wrapped by the Endoplasmic Reticulum. *Journal of Virology*, *72*(3), 2373–2387.

Rowlands, R. J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., Dwarka, R., Onashvili, T., Albina, E., & Dixon, L. K. (2008). African swine fever virus isolate, Georgia, 2007. *Emerging Infectious Diseases*, *14*(12), 1870–1874.

Salas, M. L., & Andrés, G. (2013). African swine fever virus morphogenesis. *Virus Research*, *173*(1), 29–41.

Salceda, J., Fernández, X., & Roca, J. (2006). Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA. *The EMBO Journal*, *25*(11), 2575–83.

Sanchez-Botija, C. (1963). Reservorios del virus de la peste porcina Africana. Investigación del virus de la PPA en los artrópodos mediante la prueba de la hemoadsorción. *Bull. Off. Int. Epizoot*, *60*, 895–899.

Sánchez-Vizcaíno, J. M., Mur, L., Gomez-Villamandos, J. C., & Carrasco, L. (2015). An Update on the Epidemiology and Pathology of African Swine Fever. *Journal of Comparative Pathology*, *152*(1), 9–21.

Sánchez-Vizcaíno, J. M., Mur, L., & Martínez-López, B. (2012). African swine fever: an epidemiological update. *Transboundary and Emerging Diseases, 59 Suppl 1*, 27–35.

Sander, M., & Hsieh, T. (1983). Double strand DNA cleavage by type II DNA topoisomerase from Drosophila melanogaster. *Journal of Biological Chemistry*, *258*(13), 8421–8428.

Schmidt, B. H., Osheroff, N., & Berger, J. M. (2012). Structure of a topoisomerase II-DNA-nucleotide complex reveals a new control mechanism for ATPase activity. *Nature Structural & Molecular Biology*, *19*(11), 1147–54.

Schoeffler, A. J., & Berger, J. M. (2008). DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Quarterly Reviews of Biophysics*, *41*(1), 41–101.

Schvartzman, J. B., & Stasiak, A. (2004). A topological view of the replicon. *EMBO Reports*, *5*(3), 256–61.

Shah, S., & Heddle, J. (2014). Squaring up to DNA: pentapeptide repeat proteins and DNA mimicry. *Applied Microbiology and Biotechnology*, *98*(23), 9545–9560.

Simón-Mateo, C., Andrés, G., Almazán, F., & Viñuela, E. (1997). Proteolytic processing in African swine fever virus: evidence for a new structural polyprotein, pp62. *Journal of Virology* , *71* (8 ), 5799–5804.

Simón-Mateo, C., Andrés, G., & Viñuela, E. (1993). Polyprotein processing in African swine fever virus: a novel gene expression strategy for a DNA virus. *The EMBO Journal*, *12*(7), 2977–2987.

Suarez, C., Andrés, G., Kolovou, A., Hoppe, S., Salas, M. L., Walther, P., & Locker, J. K. (2015). African swine fever virus assembles a single membrane derived from rupture of the endoplasmic reticulum. *Cellular Microbiology*, n/a–n/a.

Suárez, C., Gutiérrez-Berzal, J., Andrés, G., Salas, M. L., & Rodríguez, J. M. (2010). African Swine Fever Virus Protein p17 Is Essential for the Progression of Viral Membrane Precursors toward Icosahedral Intermediates. *Journal of Virology* , *84* (15 ), 7484–7499.

Sundin, O., & Varshavsky, A. (1981). Arrest of segregation leads to accumulation of highly intertwined catenated dimers: Dissection of the final stages of SV40 DNA replication. *Cell*, *25*(3), 659–669.

Tabarés, E., Martinez, J., Gonzalvo, F. R., & Sánchez-Botija, C. (1980). Proteins specified by African swine fever virus. *Archives of Virology*, *66*(2), 119–132.

Tabarés, E., & Sánchez Botija, C. (1979). Synthesis of DNA in cells infected with African swine fever virus. *Archives of Virology*, *61*(1-2), 49–59.

Taneja, B., Patel, A., Slesarev, A., & Mondragón, A. (2006). Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases. *The EMBO Journal*, *25*(2), 398–408.

Thomson, G. R. (1985). The epidemiology of African swine fever: the role of free-living hosts in Africa. *The Onderstepoort Journal of Veterinary Research*, *52*(3), 201–209.

Tran, J. H., Jacoby, G. A., & Hooper, D. C. (2005). Interaction of the Plasmid-Encoded Quinolone Resistance Protein Qnr with Escherichia coli DNA Gyrase. *Antimicrobial Agents and Chemotherapy*, *49*(1), 118–125.

Tse, Y. C., Kirkegaard, K., & Wang, J. C. (1980). Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *Journal of Biological Chemistry*, *255*(12), 5560–5565.

Uemura, T., Ohkura, H., Adachi, Y., Morino, K., Shiozaki, K., & Yanagida, M. (1987). DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in S. pombe. *Cell*, *50*(6), 917–925.

Vial, L., Wieland, B., Jori, F., Etter, E., Dixon, L., & Roger, F. (2007). African swine fever virus DNA in soft ticks, Senegal. *Emerging Infectious Diseases*, *13*(12), 1928–1931.

Wang, J. C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Reviews. Molecular Cell Biology*, *3*(6), 430–40.

Wang, X., Reyes-Lamothe, R., & Sherratt, D. J. (2008). Modulation of Escherichia coli sister chromosome cohesion by topoisomerase IV. *Genes & Development*, *22*(17), 2426–2433.

Wardley, R. C., Andrade, C. deM., Black, D. N., de Castro Portugal, F. L., Enjuanes, L., Hess, W. R., Mebus, C., Ordas, A., Rutili, D., Sanchez Vizcaino, J., Vigario, J. D., Wilkinson, P. J., Moura Nunes, J. F., & Thomson, G. (1983). African swine fever virus. *Archives of Virology*, *76*(2), 73–90.

Wesley, R. D., & Tuthill, A. E. (1984). Genome relatedness among African swine fever virus field isolates by restriction endonuclease analysis. *Preventive Veterinary Medicine*, *2*(1–4), 53–62.

Wessel, I., Jensen, P. B., Falck, J., Mirski, S. E. L., Cole, S. P. C., & Sehested, M. (1997). Loss of Amino Acids 1490Lys-Ser-Lys1492 in the COOH-Terminal Region of Topoisomerase IIα in Human Small Cell Lung Cancer Cells Selected for Resistance to Etoposide Results in an Extranuclear Enzyme Localization. *Cancer Research* , *57*(20 ), 4451–4454.

Wigley, D. B., Davies, G. J., Dodson, E. J., Maxwell, A., & Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature*, *351*(6328), 624–629.

Wilkinson, P. J., Pegram, R. G., Perry, B. D., Lemche, J., & Schels, H. F. (1988). The distribution of African swine fever virus isolated from *Ornithodoros moubata* in Zambia. *Epidemiology & Infection*, *101*(03), 547–564.

Witz, G., & Stasiak, A. (2010). DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic Acids Research*, *38*(7), 2119–33.

Yáñez, R. J., Rodríguez, J. M., Nogal, M. L., Yuste, L., Enríquez, C., Rodriguez, J. F., & Viñuela, E. (1995). Analysis of the complete nucleotide sequence of African swine fever virus. *Virology*, *208*(1), 249–78.

Yutin, N., Colson, P., Raoult, D., & Koonin, E. V. (2013). Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. *Virology Journal*, *10*, 106.

Zanetti, G., Pahuja, K. B., Studer, S., Shim, S., & Schekman, R. (2012). COPII and the regulation of protein sorting in mammals. *Nat Cell Biol*, *14*(1), 20–28.

Zhang, H., Barceló, J. M., Lee, B., Kohlhagen, G., Zimonjic, D. B., Popescu, N. C., & Pommier, Y. (2001). Human mitochondrial topoisomerase I. *Proceedings of the National Academy of Sciences* , *98* (19 ), 10608–10613.

Zhang, H., Meng, L. H., Zimonjic, D. B., Popescu, N. C., & Pommier, Y. (2004). Thirteen-exon-motif signature for vertebrate nuclear and mitochondrial type IB topoisomerases. *Nucleic Acids Research*, *32*(7), 2087–2092.

# CHAPTER 2

# BIOINFORMATICS AND PHYLOGENETIC ANALYSIS OF ORF P1192R OF AFRICAN SWINE FEVER VIRUS

## 2.1 Introduction

Though the era of genomics and fast genome-sequencing provides countless new nucleotide and amino acid sequences, frequently these data have no parallel support from experimental data, such as functional assays. Therefore, an easy way to start addressing the study of a gene of unknown function is through bioinformatics, and a first approach is usually a simple pairwise alignment, which may indicate similarities between sequences. The general principle of this approach is that sequences derived from a single common ancestor will have more similarities than two unrelated, randomly chosen sequences, and algorithms have been developed to assess this and confer it statistical significance [e.g. BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990)]. However, sequences may have diverged in such a way that the overall sequence similarity is low. In these cases, the search for conserved functional domains and/or motifs [e.g. InterProScan (Quevillon *et al.*, 2005)] may also provide clues on, at least, one biochemical function of a protein. If a protein sequence is found to have more than 40% of sequence identity to another protein sequence of known function, and if functionally important residues are conserved between the two sequences, this is usually evidence that both proteins share a common biochemical function. The overlying reason for this is that evolution has generated a limited number of catalytic mechanisms and of protein folds necessary to accomplish those same mechanisms. Multiple sequence alignments (MSAs) can take this approach even further since they allow the simultaneous comparison of several sequences. While in past times MSAs were usually performed manually and, therefore, comparison of a large number of sequences was unfeasible, with the improvement of computing technology and of alignment algorithms, aligning multiple sequences is somewhat trivial nowadays. Most MSAs are now constructed using the progressive sequence alignment method, by which an alignment is built stepwise, starting with the most similar sequences and progressively adding more divergent ones (Feng & Doolittle, 1987; Chenna, 2003). Commonly used packages for multiple sequence alignment construction are Clustal W (Thompson, Higgins, & Gibson, 1994), T-Coffee (Notredame, Higgins, & Heringa, 2000), MAFFT (Katoh, 2002) and MUSCLE (Edgar, 2004).

Alignments from any one of these packages can then be used, for example, in phylogenetic analyses, in which possible evolutionary relationships between sequences are inferred. The initial multiple sequence alignment is a key step in phylogenetic tree construction, since a poor quality alignment will certainly result in an erroneous phylogenetic distribution and mislead into establishing relationships that are most certainly false. Because it is impossible to trace back the exact set of events that led to the accumulation of difference between similar sequences, phylogenetic tree construction is based on evolutionary models. Each model has a

set of assumptions, more or less rigid, by which they try to explain how a certain character in a sequence (e.g. A, T, C or G) evolves from one state (for example, A) to another (G). Having determined the best model to use in a given phylogenetic analysis, the choice of method for tree construction follows. Phylogenetic methods commonly used are divided in distance-based methods – for example, the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal & Michener, 1958), the Neighbor-Joining (NJ) (Saitou & Nei, 1987) or the Minimum Evolution (ME) (Rzhetsky & Nei, 1993) methods – and in character-based methods – Maximum Parsimony (MP) (Fitch, 1971), Maximum Likelihood (ML) (Felsenstein, 1973) or Bayesian Inference (Yang & Rannala, 1997). Distance-based methods build trees by grouping sequences according to their overall dissimilarity (or distance), i.e., they simply calculate the distances between sequences based on their overall differences and then use those values to assemble the phylogenetic tree, discarding the character data from the sequences. Character-based methods determine the phylogenetic tree which optimizes the distribution of the characters as seen on the actual data (the multiple sequence alignment), thereby using the character data on all steps of the analysis.

Evaluation of a phylogenetic tree can also be performed in several ways, but the most common is through bootstrapping (Felsenstein, 1985). One of the advantages of this method is its applicability to both distance-based and character-based tree-building methods. Bootstrapping works by resampling, with replacement, the characters in the multiple sequence alignment (in other words, by resampling the data), thereby building alternative MSAs which are then used to obtain phylogenetic trees. From these multiple sets of trees, the number of times a given branch is found in the same position is counted and that number is used as a bootstrap value. These values provide a sort of probability that the true phylogeny is the one represented in the obtained phylogenetic tree, or in other words, they provide a measure of accuracy, and bootstrap values are usually considered reliable when they are equal to or higher than 70% (Hillis & Bull, 1993).

ORF P1192R was initially found in the genomic sequence of ASFV on the basis of its homology to type II DNA topoisomerases (Baylis, Dixon, Vydelingum, & Smith, 1992; García-Beato *et al.*, 1992). Although these studies included analyses with bioinformatics tools and a brief phylogenetic analysis, since then the development of resources and technologies available to perform such studies has been astounding, as has the knowledge gained on type II topoisomerases, not only on how they work, but also on how widespread they can be found among living organisms. Therefore, we sought to re-examine this ORF in the light of current knowledge.

## 2.2 Materials and methods

### 2.2.1 Bacterial strains and transformation

For cloning and plasmid amplification and purification procedures the *Escherichia coli* strain DH5α was used. Cells of this strain were made competent for transformation using an adaptation of the calcium chloride method (Mandel & Higa, 1970). For this, a 500 ml LB (NZYTech) culture was inoculated, using 5 ml of a pre-culture grown at 37 ºC overnight, and incubated at 37 ºC, 180 - 200 rpm, until the culture reached an $OD_{600}$ of about 0.4. The entire culture volume was collected into previously chilled polypropylene tubes and incubated on ice for 20 minutes. All of the procedures that followed were performed either in a cold-room or at 4 ºC. After centrifugation for 10 minutes at 3000 *g*, cell pellets were resuspended in 30 ml of cold 0.1 M $CaCl_2$ (Merck), transferred to 50 ml polypropylene tubes and further incubated on ice for 30 minutes. A 10 minute centrifugation at 3000 *g* followed, and the cell pellets were again resuspended, this time by pipetting, in 8 ml of cold 0.1 M $CaCl_2$ containing 15% (v/v) glycerol (Sigma-Aldrich). Aliquots of 150 µl of these cell suspensions were collected in 1.5 ml tubes and immediately stored at -80 ºC.

For bacterial transformation, 50 µl of thawed competent cell suspensions were mixed in 1.5 ml tubes with either 1 µl of plasmid DNA or 5 µl of ligation mixture and incubated on ice for 20 minutes. After this time, cells were subjected to a 42 ºC heat-shock for 90 seconds, followed by incubation on ice for 2 minutes. LB medium was added up to 1 ml and the tube(s) incubated at 37 ºC for 1 hour. Cells were then collected by centrifuging for 2 minutes at 7000 rpm, resuspended in 150 µl of culture medium and plated on solid LB (supplemented with 12 g/l of agar) plates supplemented with the appropriate antibiotic – ampicillin (100 µg/ml) or kanamycin (30 µg/ml) (both from NZYTech). Plates were incubated for up to 16 hours at 37 ºC to allow for bacterial colony formation.

### 2.2.2 Primers used

The primers used in this chapter are listed below in Table 1. Sequences of restriction sites included in the primers are underlined. Nucleotides included upstream of those sequences were added to increase efficiency of the hydrolysis reactions. The primers were synthesized either by Life Technologies or by STAB VIDA.

**Table 1.** Primers used in this chapter.

| Primer name | Sequence (5' → 3') | Purpose |
|---|---|---|
| P1192R ATG NdeI Up | CCCATATGATGGAAGCGTTTGAAATCAGC | cloning |
| P1192R noSTOP NotI Low | CCGCGGCCGCATGAAATTTCCACTGAGTA | cloning |
| P1192R_3homRV | GGCGCTTAGTGGTTACAAACAAAGC | cloning |
| P1192R_5homFW | CCATGATGGACTTTGAGAGGGTCC | cloning |
| P1192R_5homRV | TTTTATTCCTATAAAGAATTCAAACTTTATTAG | sequencing |
| P1192R 2nd Frag FW | CCGAATTCCCCAATTACAATTTCAAGGGCCATTTGAAGCGGTTTGGCCA | sequencing |
| P1192R iSeq1 Up | ACAAATCAACCAGCGCCT | sequencing |
| P1192R iSeq3 Up | CGACGAATTCCTTGCAGCCT | sequencing |
| P1192R iSeq4 Up | CAACAGACGATTAAAGATAAAAACC | sequencing |
| P1192R iSeq5 Up | GAGCATGGATTTCCCCCGCTG | sequencing |
| P1192R iSeq6 Low | CTATGGCAAGCTCTTTCATGATCC | sequencing |
| P1192R_seq1 | GAAGCGTTTGAAATCAGC | sequencing |
| P1192R_seq2 | TCATTGAGCCTCCTACCA | sequencing |
| P1192R_seq7 | CTTAATGGTCGCCGTATG | sequencing |
| Real-PCR_2_ FW | AGCGAGCAAAGCTGAGATACG | sequencing |
| T7_RV_JC | GCTAGTTATTGCTCAGCGG | sequencing |
| T7_FW_JC | TAATACGACTCACTATAGGG | sequencing |

Sequences recognized by restriction enzymes are shown underlined – CATATG for *Nde*I, GCGGCCGC for *Not*I and GAATTC for *Eco*RI. Some of the primers used for sequencing were also used for colony PCR.

### 2.2.3 DNA amplification by PCR for cloning purposes

To generate DNA fragments that could be used for cloning, PCR reactions were performed using a Mastercycler gradient thermal cycler (Eppendorf) and a 100 µl volume in 200 µl microtubes. For proofreading purposes, enzyme Phusion High-Fidelity DNA polymerase (Thermo Scientific) was used, at 0.02 U/µl, together with Phusion HF Buffer. A solution of deoxynucleotides (dNTPs; Thermo Scientific), containing dATP, dTTP, dGTP and dCTP, was added at a final concentration of 200 µM, as were the forward and reverse primers at a final concentration of 0.5 µM each. The conditions used were the following: tubes were heated to 98 ºC for 2 minutes to allow for DNA denaturation, and this was followed by 30 cycles of denaturation at 98 ºC for 30 seconds, annealing at the optimal annealing temperature for 30 seconds, and extension at 72 ºC for 15 to 30 seconds per 1000 base-pairs of amplicon. A final extension step of 10 minutes at 72 ºC concluded the PCR conditions. The optimal annealing temperature was determined for each pair of primers by performing a gradient PCR using the conditions described above but with reactions of 20 µl of volume. Extension at 72 ºC for 15 seconds per 1000 base-pairs was used when low complexity DNA

(plasmid) was used as template, while an extension time of 30 seconds per 1000 base-pairs was used when the DNA template was of high complexity (genomic DNA).

Frequently, several identical PCR reactions were prepared for amplification of a specific fragment. After confirmation of successful amplification, reactions were pooled in a single tube, to which 0.1 volumes of 3 M sodium acetate (Sigma-Aldrich), pH 5.2, and 2.5 volumes of ice-cold absolute ethanol (Merck) were added. The tube was incubated at least overnight at -20 ºC to promote DNA precipitation. This was followed by centrifugation at 13200 rpm for 15 minutes, to collect the precipitated DNA. The supernatant was discarded, the pellet washed with 750 μl of 70% (v/v) ethanol and the tube was centrifuged again for 5 minutes at 13200 rpm. After discarding the supernatant, the pellet was air-dried and resuspended in a small volume of Milli-Q water (filtered and deionized, with resistivity ≥ 18 MΩ).

### 2.2.4 Agarose gel electrophoresis

The percentage of agarose (Lonza) in the gels used for analysis of DNA fragments depended on the size of the latter and usually varied between 0.7 and 1.5% (w/v). Electrophoresis was performed using TAE buffer (40 mM Tris-Acetate, 1mM EDTA) both for the preparation of the gel and of the running buffer. For nucleic acid staining, the gel was either stained for 30 minutes in a 0.5 μg/ml ethidum bromide (Sigma-Aldrich) solution after running, or 1% (v/v) of RedSafe (iNtRON Biotechnology) was added to allow for nucleic acid staining during the run (in this case, RedSafe was usually excluded from the TAE buffer). Before being loaded, samples were mixed with DNA loading buffer [50% (v/v) glycerol, 0.45% (w/v) bromophenol blue], except those resulting from colony PCRs which already included loading dye in the reaction buffer. Gels were usually run at 90 - 120 V for 30 - 60 minutes, followed by visualization in a transiluminator upon exposure to ultraviolet light. Determination of the size of DNA fragments was done by comparison with DNA ladders HyperLadder I (Bioline) or NZYDNA Ladder III (NZYTech).

### 2.2.5 DNA purification from agarose slices

Purification of DNA fragments originating from PCRs or from hydrolyses with restriction enzymes was performed using the Zymoclean Gel DNA Recovery kit (ZymoResearch). Briefly, after agarose gel electrophoresis the bands with the expected size and predicted to contain the desired DNA were excised from the gel with a blade, placed in microcentrifuge tubes and the Agarose Dissolving Buffer was added as indicated by the manufacturer. The agarose slices were then incubated at 55 ºC until complete dissolution, and the solutions were

placed onto columns supplied in the kit. After washing twice with the supplied DNA Wash Buffer, DNA from each column was eluted using 20 µl of Milli-Q water.

### 2.2.6 Plasmid DNA purification

Purification of recombinant DNA molecules was done using the commercial kits QIAprep Miniprep kit (QIAGEN), NZYMidiprep kit (NZYTech) or GenElute HP Plasmid Maxiprep kit (Sigma-Aldrich), according to the manufacturers' instructions. Sequencing of DNA for confirmation of its sequence was performed by STAB VIDA.

### 2.2.7 DNA quantification

Concentration and purity of DNA molecules were determined by using NanoDrop 2000c spectrophotometer (Thermo Scientific), under ultraviolet light of 260 nm and 280 nm. For the concentration it was assumed that an optical density unit at 260 nm corresponds to a double-stranded DNA concentration of 50 µg/ml. Purity was evaluated using the ratio between optical densities at 260 nm and 280 nm.

### 2.2.8 Hydrolysis and ligation of DNA fragments

Restriction enzymes used for hydrolysis of plasmids and/or DNA fragments were *Nde*I and *Not*I (both from New England Biolabs). Usually, around 2.5 µg of plasmid DNA or about 5.0 µg of a DNA fragment were hydrolyzed by incubating them overnight at 37 ºC with 10 - 20 units of each enzyme, in the appropriate reaction buffer according to the enzyme manufacturer. For reactions with more than one restriction enzyme, the buffer with the best efficiency for both enzymes was chosen. When necessary, these conditions were adjusted to obtain the desired results.

Ligation reactions between plasmids and DNA fragments were prepared in a total of 20 µl of reaction volume, using 5 - 10 µl of fragment solution, 1 µl of plasmid solution, 2 µl of 10x T4 DNA Ligase Buffer, 5 U of T4 DNA Ligase (Thermo Scientific) and $H_2O$ to complete the volume. Reactions were incubated at room temperature for 1 hour, after which they were incubated overnight at 4 ºC.

### 2.2.9 Cloning and sequencing of P1192R from ASFV/L60 and ASFV/E75

L60 genomic DNA was obtained from infected pig macrophages and isolated using a High Pure Viral Nucleic Acid Kit (Roche). The complete ORF P1192R (skipping the endogenous stop codon) was PCR-amplified using specific primers including *Nde*I and *Not*I restriction

sites and cloned in *Nde*I and *Not*I sites of the pET24a expression vector (Novagen). These primers were designed based on the information available for ORF P1192R of isolate Ba71V.

Genomic DNA of isolate ASFV/E75 (E75) (de Villiers *et al.*, 2010), used for sequencing of ORF P1192R, was kindly provided by Dr. Carmina Gallardo (CISA-INIA, Spain). A fragment starting 519 bp upstream and ending 511 bp downstream of ORF P1192R was PCR-amplified from E75 genomic DNA using specific primers and cloned in the pJET1.2/blunt cloning vector (Thermo Scientific).

Sequencing was performed using primers indicated on Table 1. These primers were designed considering that sequencing readings would contain useful information from about 800 nucleotides, and in order to obtain overlaps between readings of 100 - 200 nucleotides.

### 2.2.10 Colony PCR

Bacterial colonies resulting from transformations for cloning purposes were screened for positives by PCR (colony PCR) using specific primers for each construct. Since proofreading wasn't required, DreamTaq DNA Polymerase (Thermo Scientific) was used in this case. Each colony was selected using a micropipette tip, streaked on an LB plate (supplemented with the appropriate antibiotic) and dipped in a PCR tube containing 20 µl of colony PCR mixture (1X DreamTaq Buffer Green, 200 µM dNTPs, 0.5 µM of each primer, 1.25 U of DreamTaq DNA Polymerase). After resuspending cells in the mixture, the tubes were placed on a Mastercycler gradient thermal cycler and subjected to the following conditions: tubes were heated to 95 ºC for 3 minutes followed by 25 cycles of denaturation at 95 ºC for 30 seconds, annealing 55 ºC for 30 seconds and extension at 72 ºC for 60 seconds per 1000 base-pairs of amplicon. A final extension step of 10 minutes at 72 ºC concluded the PCR conditions. The primers used were chosen to generate amplicons of similar sizes, so that the conditions used remained constant.

### 2.2.11 Bioinformatics analyses

Definition of domains in type II topoisomerases was performed using the InterProScan resource available at http://www.ebi.ac.uk/Tools/pfa/iprscan/. Search for conserved motifs in the P1192R deduced amino acid sequence was carried out with the use of patmatdb and pairwise alignments of type II topoisomerases were done using needle, both of which are EMBOSS tools (Rice, Longden, & Bleasby, 2000). The presence of nuclear localization signals was predicted using the PSORT II program (Nakai & Horton, 1999) and NetNES v1.1 (la Cour *et al.*, 2004) was used for nuclear export signal prediction. Prediction of subcellular localization sites was done using CELLO (Yu, Chen, Lu, & Hwang, 2006) and PSORT II, and Phobius (Käll, Krogh, & Sonnhammer, 2007) and SVMtm (Yuan, Mattick, & Teasdale, 2004)

were used for analysis of the presence of transmembrane domains. Modeling of the tertiary structure of pP1192R was performed using the RaptorX structure prediction web-server (http://raptorx.uchicago.edu/StructurePrediction/predict/) (Källberg *et al.*, 2012).

### 2.2.12 Sequence alignment and phylogenetic analyses

DNA sequences of ORF P1192R were obtained from sequencing (L60 and E75), from personal communication by R. Portugal and G. Keil (ASFV/NH/P68) or retrieved from ASFV genomic sequences available at the Asfarviridae Bioinformatics Resource (http://athena.bioc.uvic.ca/organisms/Asfarviridae) or from GenBank (Ba71V, NC_001659.1; ASFV/Benin97/1, AM712239.1; ASFV/OURT88/3, AM712240.1; E75, FN557520.1; ASFV/Mkuzi1979, AY261362.1; ASFV/Georgia2007/1, FR682468.1; ASFV/Tengani62, AY261364.1; ASFV/Pretorisuskop96/4, AY261363.1; ASFV/Warthog, AY261366.1; ASFV/Warmbaths, AY261365.1; ASFV/Malawi Lil/20/1, AY261361.1; ASFV/Kenya1950, AY261360.1). Their respective amino acid sequences were obtained through *in silico* translation of the nucleotide sequences and used in protein-protein BLAST searches (with a threshold of 0.0001 and remaining options as default) performed against the UniProtKB database (www.uniprot.org). Protein sequences obtained from BLAST, as well as other sequences obtained from ontology searches, and used for multiple sequence alignments are indicated in Table 2. For comparison, the sequences of the three subunits of the phage T4 type II topoisomerase were merged in a single sequence, as were the sequences of DNA gyrase subunits of *E. coli* or *Mycobacterium tuberculosis*. Alignments of amino acid sequences were performed using M-COFFEE (Moretti *et al.*, 2007) while DNA sequence alignments were done using CLUSTALW2 (Larkin *et al.*, 2007) at http://www.ebi.ac.uk/Tools/msa/clustalw2/ (Goujon *et al.*, 2010), both with default parameters. TrimAl (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) was used for protein alignment curation, with most default settings but choosing a gap threshold (fraction of positions without gaps in a column) of 0.8 and a minimum percentage of positions to conserve of 20. ProtTest 3.0 (Darriba, Taboada, Doallo, & Posada, 2012) and jModelTest 2.1.2 (Darriba, Taboada, Doallo, & Posada, 2012) were used to select the best model for phylogenetic tree construction based on protein and DNA alignments, respectively. Maximum likelihood (ML) trees were constructed using PhyML 3.0 (Guindon *et al.*, 2010) with 1000 or 100 (when 1000 was impossible to achieve due to computational limitations) bootstrap replicates, using settings indicated by previous determination of the best phylogenetic model. Trees were edited using the program MEGA 5 (Tamura *et al.*, 2011).

**Table 2.** Topoisomerase sequences used for phylogenetic analyses

| GenBank accession no. | Definition / Sequence name | Collapsed tree branch |
|---|---|---|
| BAA02076 | DNA topoisomerase II [*Mus musculus*] | Top2alpha Vertebrates |
| CAA86496 | DNA topoisomerase II [*Rattus norvegicus*] | |
| AAA37023 | DNA topoisomerase II [*Cricetulus griseus*] | |
| AAC77388 | topoisomerase II alpha [*Homo sapiens*] | |
| BAA23778 | topoisomerase II [*Sus scrofa*] | |
| BAA22539 | DNA topoisomerase II alpha [*Gallus gallus*] | |
| AAT68150 | topoisomerase 2 [*Danio rerio*] | |
| AAH41106 | DNA topoisomerase II beta [*Mus musculus*] | Top2beta Vertebrates |
| BAE96971 | DNA topoisomerase II beta [*Rattus norvegicus*] | |
| CAA48197 | DNA topoisomerase II beta [*Homo sapiens*] | |
| BAA22540 | DNA topoisomerase II beta [*Gallus gallus*] | |
| BX927238 | Zebrafish DNA sequence, clone DKEY-98N4 in linkage group 19, complete seq. | |
| AAB67168 | topoisomerase II [*Bombyx mori*] | Top2 Insects |
| CAA43523 | DNA topoisomerase type II [*Drosophila melanogaster*] | |
| CAA88867 | Protein TOP-2 [*Caenorhabditis elegans*] | n. c. |
| Q9Y8G8 | DNA topoisomerase II [*Penicillium chrysogenum*] | Top2 Fungi |
| CAA20107 | DNA topoisomerase II [*Schizosaccharomyces pombe*] | |
| CAA71405 | topoisomerase II [*Candida albicans*] | |
| AAB36610 | topoisomerase II [*Saccharomyces cerevisiae*] | |
| BAA33955 | topoisomerase II [*Candida glabrata*] | |
| AAN85207 | DNA topoisomerase II [*Nicotiana tabacum*] | Top2 Plants |
| AAA65448 | topoisomerase II [*Arabidopsis thaliana*] | |
| CAA74891 | topoisomerase II [*Pisum sativum*] | |

**Table 2** (continuation)

| | | |
|---|---|---|
| EAL72559 | DNA topoisomerase II [*Dictyostelium discoideum* AX4] | n. c. |
| X79345 | Plasmodium falciparum TopoII gene [*Plasmodium falciparum*] | n. c. |
| AAS90119 | DNA topoisomerase type 2 A [*Tetrahymena thermophila*] | Top2 T. termophila |
| AAS90120 | DNA topoisomerase type 2 B [*Tetrahymena thermophila*] | |
| CCD73199 | Protein R05D3.12, isoform a [*Caenorhabditis elegans*] | n. c. |
| AAV50746 | topoisomerase II [*Acanthamoeba polyphaga* mimivirus] | n. c. |
| ADO67359 | putative DNA topoisomerase IIA [*Cafeteria roenbergensis* virus BV-PW1] | n. c. |
| AGM15363 | DNA gyrase/DNA topoisomerase IV subunit A [*Phaeocystis globosa* virus] | n. c. |
| ADX05889 | putative DNA gyrase/DNA topo IV subunit A [Organic Lake phycodnavirus 1] | n. c. |
| ABT16685 | hypothetical protein ATCV1_Z551L [*Acanthocystis turfacea Chlorella* virus 1] | Top2 Chlorovirus |
| AGE59835 | DNA topoisomerase II [*Acanthocystis turfacea Chlorella* virus TN603.4.2] | |
| AAU95770 | topoisomerase II [*Paramecium bursaria Chlorella* virus Marburg 1] | |
| AAC96932 | DNA topoisomerase II [*Paramecium bursaria Chlorella* virus 1] | |
| ADQ91360 | hypothetical protein BpV2_193c [*Bathycoccus* sp. RCC1105 virus BpV2] | Top2 Prasinovirus |
| ADQ91815 | hypothetical protein BpV1_188c [*Bathycoccus* sp. RCC1105 virus BpV1] | |
| CBI70210 | DNA topoisomerase II [*Ostreococcus tauri* virus 2] | |
| CAY39800 | DNA topoisomerase II [*Ostreococcus tauri* virus 1] | |
| ABF82116 | hypothetical protein MIV086L [*Aedes taeniorhynchus* iridescent virus] | Top2 Iridovirus |
| AAD48151 | 045L [Invertebrate iridescent virus 6] | |
| ADB03954 | DNA topoisomerase II [Marseillevirus marseillevirus] | n. c. |
| AAA85575 | DNA topoisomerase II [*Trypanosoma cruzi*] | Top2 Trypanosomatidae |
| AAA30256 | DNA topoisomerase II [*Trypanosoma brucei*] | |
| AAC05295 | DNA topoisomerase II [*Leishmania chagasi*] | |
| CAA42182 | kinetoplast-associated type II DNA topoisomerase (toptIImt) [*Crithidia fasciculata*] | |

**Table 2** (continuation)

| | | |
|---|---|---|
| AAD42506 | gp60 topoisomerase II, large subunit, C-terminal region [Enterobacteria phage T4] | fused in a single sequence |
| AAD42487 | gp52 topoisomerase II, medium subunit [Enterobacteria phage T4] | |
| AAD42462 | gp39 topoisomerase II, large subunit, N-terminal region [Enterobacteria phage T4] | |
| BAA20341 | DNA gyrase subunit B [*Escherichia coli* str. K12] | fused in a single seq. |
| AAC75291 | DNA gyrase (type II topoisomerase), subunit A [*Escherichia coli* str. K-12] | |
| CCP42727 | DNA gyrase (subunit B) GyrB [*Mycobacterium tuberculosis* H37Rv] | fused in a single seq. |
| AAA83017 | gyrase A [*Mycobacterium tuberculosis* H37Rv] | |
| EAL67548 | DNA topoisomerase II, mitochondrial [*Dictyostelium discoideum* AX4] | n. c. |
| AAP83584 | DNA-topoisomerase II [*Physarum polycephalum*] | n. c. |
| CAA28164 | DNA topoisomerase I TopA [*Escherichia coli*] | n. c. |
| CCP46469 | DNA topoisomerase I TopA [*Mycobacterium tuberculosis* H37Rv] | n. c. |
| AAA61206 | topoisomerase I [*Homo sapiens*] | n. c. |
| CAA99005 | TOP1 [*Saccharomyces cerevisiae*] | n. c. |
| AAA02841 | topoisomerase I [Vaccinia virus] | n. c. |
| CAA47588 | DNA topoisomerase [Variola virus] | n. c. |

Non-collapsed sequences are indicated as n. c.

## 2.3 Results

### 2.3.1 Determination of the sequence from L60's ORF P1192R and its analysis using bioinformatics tools

After being cloned in the expression vector pET24a, the sequence of P1192R from ASFV isolate L60 was determined using P1192R sequencing primers listed on Table 1. The obtained nucleotide sequence as well as the predicted amino acid sequence are presented on Figure 9. The nucleotide sequence was deposited on GenBank, under accession KM261522.

A statistical analysis of the nucleotide sequence indicated that this ORF, composed of 3579 nucleotides, has a GC content of 47.25%, which contrasts with the 38.61% found for the whole L60 genome or the 40.07% found for the concatenation of the nucleotide sequences of the annotated ORFs of the L60 genome. To determine if the considerably higher GC content was exclusive of P1192R or if it would extend to other components of the replicational or transcriptional machinery, we analyzed other ASFV ORFs (indicated on Table 3). These included G1211R, coding for the viral DNA polymerase, with a GC content of 43.48%; E301R, which putatively codes for a viral homolog of the proliferating cell nuclear antigen (PCNA), with 42.83%; C962R, an ORF coding for a putative primase, with 46.69%; NP1450L, which codes for the main subunit of the viral RNA polymerase, with 47.28%; F1055L, coding for a putative replication origin-binding helicase protein, with 38.10%; ORF M1249L, the biggest ASFV unassigned open reading frame, with 43.29%; I215L, which codes for a viral ubiquitin E2-ligase protein, with 38.43%; B646L, coding for vp72, an abundant structural protein, with 44.31%; and EP402R, an ORF coding for a CD2 homolog, having 30.57% of GC content. Most of these ORFs were found to have GC content higher than that of the entire genome. However, this was not exclusive of ORFs coding for proteins indicated as being involved in replicative and/or transcriptional processes, as an unassigned gene, M1249L, as well as a gene coding for an abundant structural protein, B646L, also have higher GC content than the entire genome.

We were then interested in knowing if the high GC content was a characteristic of ORFs coding for type II topoisomerases. For this we compared the GC content of the topoII ORF and of the genome in *S. cerevisiae*, *Acanthocystis turfacea Chlorella* virus 1 (ATCV-1) and invertebrate iridescent virus 6 (IIV-6). While the GC content of P1192R was 1.22 times that of the ASFV genome, *TOP2*'s GC content was 0.95 times that of yeast's genome, ATCV-1's was 1.07 and IIV-6's was 1.02 times. Therefore, even though no clear pattern attributable to topoII-coding ORFs was found, a GC content higher than the genome's is observed for viral ORFs coding for type II topoisomerases.

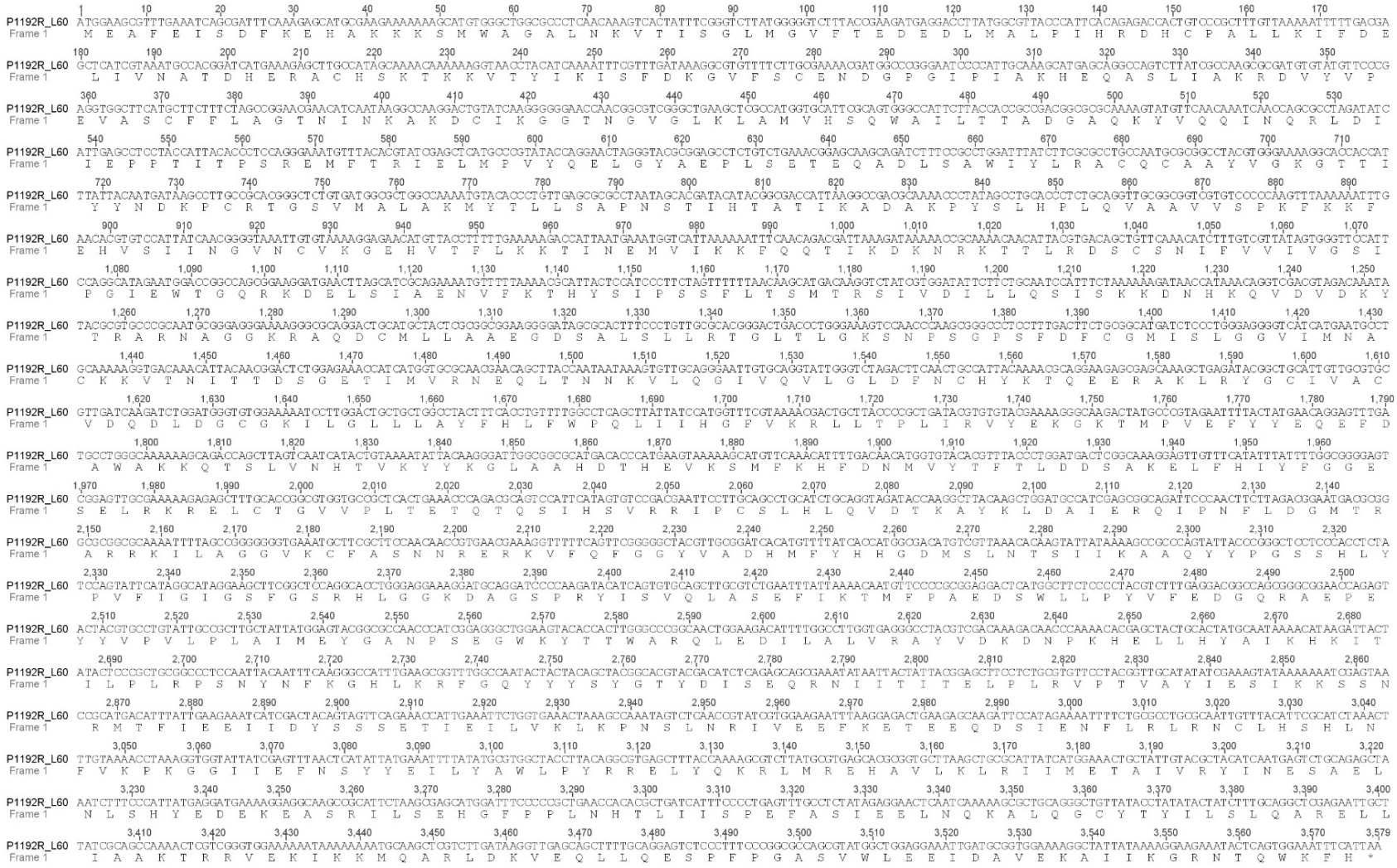


**Figure 9.** ORF P1192R from ASFV isolate L60.

Represented are the nucleotide sequence, obtained from sequencing, as well as its respective amino acid sequence obtained from *in silico* translation.

**Table 3.** GC content of P1192R and other ORFs.

| | GC content (%) | VS genome | VS concat. ORFs |
|---|---|---|---|
| L60 P1192R (type II topoisomerase) | 47.25 | 1.224 | 1.179 |
| L60 genome | 38.61 | | |
| L60 concatenated ORFs | 40.07 | | |
| L60 G1211R (DNApol) | 43.48 | 1.126 | 1.085 |
| L60 E301R (PCNA-like) | 42.83 | 1.109 | 1.069 |
| L60 C962R (primase) | 46.69 | 1.209 | 1.165 |
| L60 NP1450L (RNApol1) | 47.28 | 1.225 | 1.180 |
| L60 F1055L (origin-binding protein) | 38.10 | 0.987 | 0.951 |
| L60 M1249L (unassigned) | 43.29 | 1.121 | 1.080 |
| L60 I215L (ubiquitin E2-ligase) | 38.43 | 0.995 | 0.959 |
| L60 B646L (vp72) | 44.31 | 1.148 | 1.106 |
| L60 EP402R (CD2 homolog) | 30.57 | 0.792 | 0.763 |
| *S. cerevisiae TOP2* (type II topoisomerase) | 36.04 | 0.948 | |
| *S. cerevisiae* genome | 38.00 | | |
| ATCV-1 Z551L (type II topoisomerase) | 52.82 | 1.069 | |
| ATCV-1 genome | 49.40 | | |
| IIV-6 045L (type II topoisomerase) | 29.07 | 1.015 | |
| IIV-6 genome | 28.63 | | |

The GC content is displayed in terms of percentage of the total nucleotides per ORF or genome. The VS columns display relative values of the respective ORF in relation to its corresponding genome or concatenation of ORFs (in the case of ASFV). Grey shading highlights ORFs in which the GC content is lower than that of the respective genome.

By translating *in silico* the sequence of P1192R we obtained a protein sequence of 1192 amino acids, corresponding to a protein with a predicted molecular weight of about 135.5 kDa and with a predicted pI of 8.65. We also used the predicted amino acid sequence of P1192R to run algorithms (on webservers) for prediction of the sub-cellular protein localization, namely PSORT II, which indicated a 56.5% probability of it being cytoplasmic, together with 17.4% and 13.0% probabilities of it being nuclear or cytoskeleton-associated, respectively, and CELLO, which indicated P1192R to be cytoplasmic with a reliability of 1.624, even though nuclear localization was presented with a reliability of 1.144. As comparison, we ran the same algorithms for the sequence of the yeast type II DNA topoisomerase, Top2p, known to be a nuclear protein, and we obtained an 82.6% probability for nuclear localization in PSORT II

and a 4.210 reliability value in CELLO. For P1192R we also obtained a prediction for a putative NLS at positions 293 - 299 (PKFKKFE), using PSORT II, as well as a putative NES at residues 443 - 452 (LSLLRTGLTL), predicted by NetNES 1.1. One of our preliminary bioinformatics analyses indicated the presence of a transmembrane domain, covering positions 552 - 569. To confirm this, we used the SVMtm and Phobius algorithms, and while the former indicated a transmembrane segment encompassing residues 549 - 566, the latter predicted a transmembrane segment at positions 548-568, with the N-terminal region being non-cytoplasmic (Figure 10). Thus, both algorithms predicted the existence of a transmembrane domain in the same region of the protein sequence, in agreement with the initial prediction. Again, for comparison, the sequence of Top2p was run in SVMtm and Phobius and no transmembrane domains were detected.

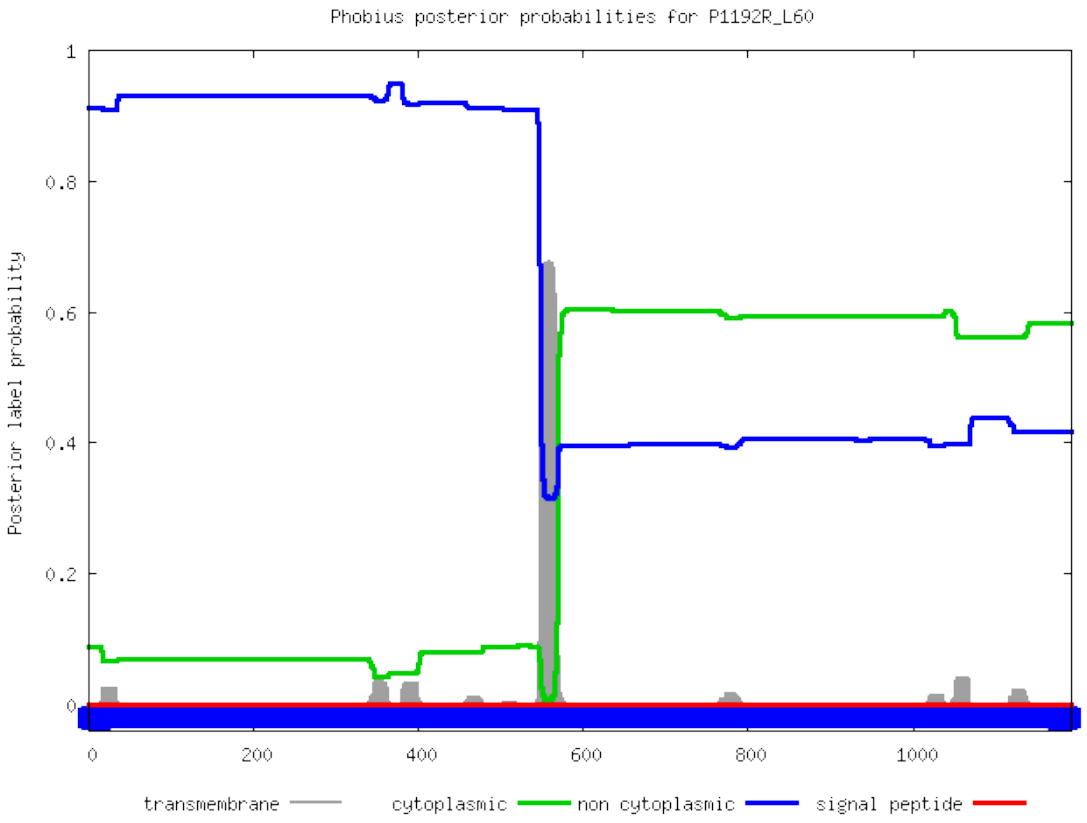


**Figure 10.** Phobius prediction of transmembrane domains in P1192R amino acid sequence.

Eukaryotic type II topoisomerases are typically composed of an N-terminal ATP-binding domain, a central catalytic domain and a C-terminal regulatory domain (Champoux, 2001). Our bioinformatic analysis on the deduced amino acid sequence showed that the depicted organization is fairly conserved, but in this particular case the C-terminal domain is missing, with the central domain extending up to the end of the protein sequence (Figure 11).

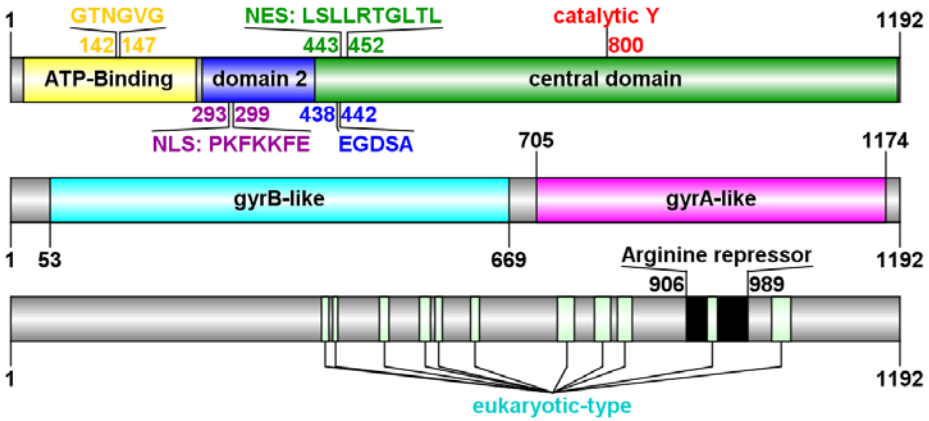


**Figure 11.** Functional motifs and domains in the amino acid sequence of P1192R.
The upper and lower schemes indicate the type II DNA topoisomerase specific domains found in the predicted amino acid sequence of P1192R, as well as potential functional and regulatory motifs. These include the ATP-binding site (GTNGVG at positions 142-147), the quinolone-resistance determining region (EGDSA at positions 438-442) and the catalytic residue (Y at position 800), as well as putative NLS (PKFKKFE at positions 293-299) and NES (LSLLRTGLTL at positions 443-452) motifs. The middle scheme indicates domains of P1192R homologous to its bacterial counterparts gyraseB (N-terminal) and gyraseA (C-terminal).

Searching the P1192R amino acid sequence for motifs which have been described as being highly conserved among type II topoisomerases, such as RPXXYIGS, GXGXP, GGXXGXG, EGDSA, PL(R/K)GK(I/L/M)LNV, IM(T/A)D(Q/A)DXD and YKGLG (Austin and Marsh, 1998), we found only matches for some, namely GPGIP, GGTNGVG (which includes the predicted ATP-binding site, shown underlined) and EGDSA. In addition, using PRINTS to search for conserved motifs we found nine N-terminal conserved motifs, in which the above mentioned three are included, and eleven motifs recognized as “eukaryotic-type”, distributed throughout the central domain of the protein and aligning extremely well with those identified through the same method for other type II topoisomerases, like Top2p, topoIIα from *Homo sapiens* or topoII from the *Paramecium bursaria Chlorella* virus Marburg 1. The predicted catalytic residue for P1192R is a tyrosine at position 800, contained in the motif SPRYI which is conserved among eukaryotic type II topoisomerases.

We attempted to predict the tertiary structure of pP1192R, and for this we used the RaptorX structure predicition web-server. We provided the predicted amino acid sequence of pP1192R and the server, based on its algorithm, accounted for both sequence and secondary structure information to search for a template from which the tertiary structure of pP1192R could be modelled. The best template, as defined by a template structural alignment, was that of *S. cerevisiae*’s Top2p, truncated at position 1177 (therefore lacking its C-terminal region;

PDB: 4gfhA) (Schmidt, Osheroff, & Berger, 2012). The obtained tertiary structure for the topoisomerase II dimer of pP1192R is presented on Figure 12.

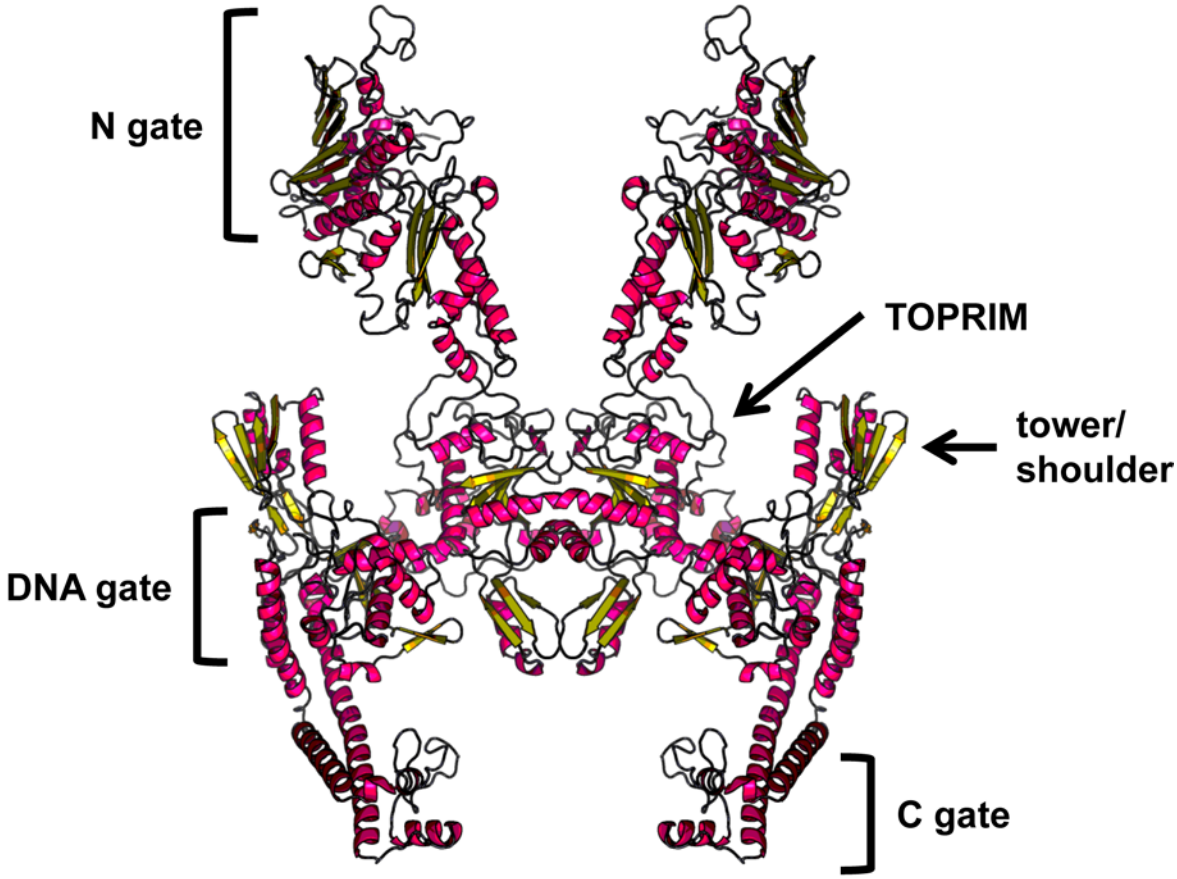


**Figure 12.** Predicted structure of pP1192R, in its dimer form.
The tertiary structure of pP1192R was predicted using the RaptorX structure prediction web-server and modeled by comparison with that of *S. cerevisiae*'s Top2p (PDB: 4gfhA). For comparison purposes (with, for example, Figure 5 – see Introduction), the structure obtained was duplicated, flipped horizontally and overlapped with the original, to produce a dimer of pP1192R, mimicking that active form of the ASFV topoisomerase II.

From the predicted structure of pP1192R it is possible to observe that the overall quaternary structure of the dimer is similar to what has been determined for type II topoisomerases (Berger, Gamblin, Harrison, & Wang, 1996; Morais Cabral *et al.*, 1997; Fass, Bogden, & Berger, 1999; Classen, Olland, & Berger, 2003; Dong & Berger, 2007; Tretter, Schoeffler, Weisfield, & Berger, 2010; Schmidt *et al.*, 2012). The structure of the N gate, which includes the ATP-binding domain, is very similar, even if in the pP1192R structure the two subunits are further apart than in determined crystal structures. The tower/shoulder domain and the C gate, including its characteristic C-terminal globular domain, are also distinguishable. However, there are also some differences to known structures, and these lie mainly in the TOPRIM/DNA gate domains, which present a slightly different conformation.

Of note is an extruding portion in the catalytic pocket of the DNA gate, not seen in crystal structures of type II DNA topoisomerases.

### 2.3.2 Phylogenetic analyses of P1192R among ASFV isolates

Several ASFV isolates of different virulence and geographic origin have had their genome fully sequenced. Therefore, we compared the nucleotide sequence of ORF P1192R from L60 to that of thirteen other ASFV isolates. In the analyzed genomes ORF P1192R is composed by 3579 nucleotides, with the E75 isolate being the exception, in which the annotated P1192R ORF has only 3434 nucleotides (de Villiers *et al.*, 2010). When looking at the nucleotide sequence of P1192R in E75, a nucleotide is missing at position 2176 (numbering relates to the other 13 P1192R sequences), which leads to a frameshift. This frameshift induces the appearance of a premature stop codon at positions 2176-2178, which translates into a protein with 725 amino acids. At position 3443, the E75 sequence also contains an extra nucleotide, which restores the original frame of the ORF *via* another frameshift. If the first frameshift were to be skipped, this second frameshift would originate a stop codon at positions 3457 - 3459, thus translating to a protein with 1152 amino acids. However, according to the annotation for P1192R or the information in the UniProtKB database, the P1192R protein from the E75 isolate is described as having 1145 amino acids in length, which isn't in accordance to any of the possible early stop codons. Since the frameshifts mentioned above originate in stretches of G's and A's, respectively, and because at positions 1491 and 1492 of the E75 sequence two N's can be found, we favor the hypothesis that the E75 sequence contains errors. To clarify this, we cloned ORF P1192R from E75, sequenced it and found it to be identical to the sequence obtained for L60 and to the sequence available for Ba71V. Using the corrected E75 sequence, we aligned the fourteen P1192R DNA sequences and used this alignment to perform a maximum likelihood phylogenetic analysis. The distribution of the isolates in the phylogenetic tree obtained (Figure 13A) strongly resembles that of the vp72 genotyping (Gallardo *et al.*, 2009). The isolates are separated into three clusters, with one main cluster including the five isolates from the Iberian Peninsula and the Benin97/1 from western Africa, together with one isolate from South Africa (Mkuzi1979), while another cluster consisted of four isolates from the eastern (Tengani62) and southern (Pretorisuskop96/4; Warmbaths; Warthog) African regions together with the recently described Georgia2007/1 isolate, which curiously branched in this latter cluster in opposition to what happens in the vp72 genotype tree. A third cluster contained the Kenya1950 and Malawi Lil/20/1 isolates, derived from the eastern part of Africa, and given the genetic differences of the isolates this cluster ended up working as an outgroup.

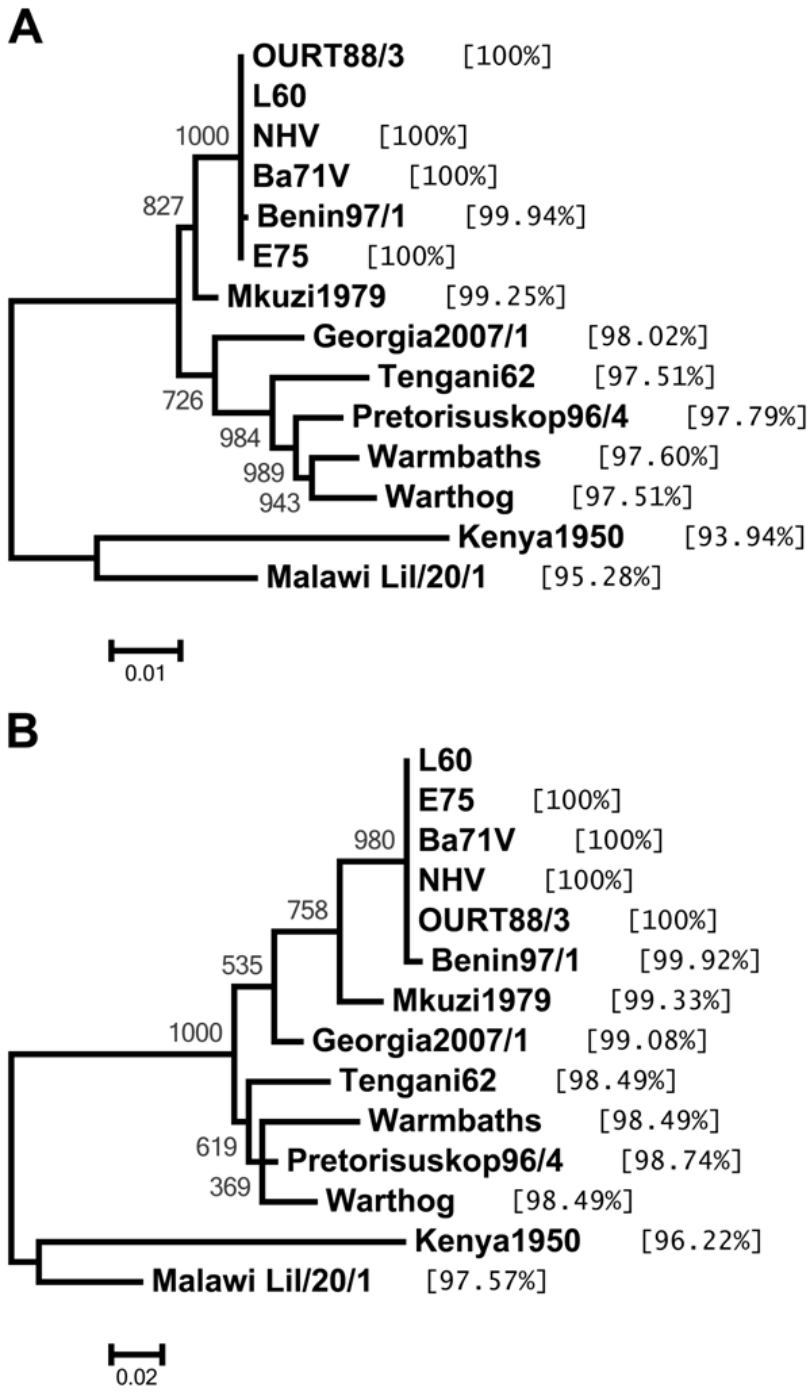


**Figure 13.** Phylogenetic analysis of ASFV P1192R.
(A) Maximum-likelihood phylogenetic tree constructed from a multiple nucleotide sequence alignment of P1192R from 14 different ASFV isolates. (B) Maximum-likelihood phylogenetic tree constructed from a multiple amino acid sequence alignment of P1192R from 14 different ASFV isolates. Values indicated between square brackets refer to percentage identity of the respective sequence to the L60 sequence. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

An analysis of the deduced P1192R amino acid sequences from the various isolates resulted in a similar tree topology (Figure 13B) with minor differences. The Pretorisuskop96/4 isolate is not clearly separated from the Warmbaths and Warthog isolates, and the Georgian isolate is now branching closer to the genotype I isolates as seen for vp72 genotyping.

Both phylogenetic trees reflect what is observed in terms of identity among the 14 sequences analyzed (see Table 4). While sequences from the European isolates are 100% identical and highly similar to the West African Benin97/1 isolate, variation is observed among isolates from the southern and eastern regions of Africa.

**Table 4.** DNA and amino acid identity percentages among P1192R from fourteen ASFV isolates.

| Isolate | | L60 | NHV | E75 | Ba71V | OUR | Benin | Mkuzi | Georgia | Tengani | Pretoris | Warm | Warthog | Kenya | Malawi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GenBank accession no. | | KM261522 | | FN557520.1 | NC_001659.1 | AM712240.1 | AM712239.1 | AY261362.1 | FR682468.1 | AY261364.1 | AY261363.1 | AY261365.1 | AY261366.1 | AY261360.1 | AY261361.1 |
| Origin | | Portugal | Portugal | Spain | Spain | Portugal | Benin | South Africa | Georgia | Malawi | South Africa | South Africa | Namibia | Kenya | Malawi |
| L60 | DNA | - | 100% | 100% | 100% | 100% | 99,94% | 99,25% | 98,02% | 97,51% | 97,79% | 97,60% | 97,51% | 93,94% | 95,28% |
| | Protein | - | 100% | 100% | 100% | 100% | 99,92% | 99,33% | 99,08% | 98,49% | 98,74% | 98,49% | 98,49% | 96,22% | 97,57% |
| NHV | DNA | 100% | - | 100% | 100% | 100% | 99,94% | 99,25% | 98,02% | 97,51% | 97,79% | 97,60% | 97,51% | 93,94% | 95,28% |
| | Protein | 100% | - | 100% | 100% | 100% | 99,92% | 99,33% | 99,08% | 98,49% | 98,74% | 98,49% | 98,49% | 96,22% | 97,57% |
| E75 | DNA | 100% | 100% | - | 100% | 100% | 99,94% | 99,25% | 98,02% | 97,51% | 97,79% | 97,60% | 97,51% | 93,94% | 95,28% |
| | Protein | 100% | 100% | - | 100% | 100% | 99,92% | 99,33% | 99,08% | 98,49% | 98,74% | 98,49% | 98,49% | 96,22% | 97,57% |
| Ba71V | DNA | 100% | 100% | 100% | - | 100% | 99,94% | 99,25% | 98,02% | 97,51% | 97,79% | 97,60% | 97,51% | 93,94% | 95,28% |
| | Protein | 100% | 100% | 100% | - | 100% | 99,92% | 99,33% | 99,08% | 98,49% | 98,74% | 98,49% | 98,49% | 96,22% | 97,57% |
| OUR | DNA | 100% | 100% | 100% | 100% | - | 99,94% | 99,25% | 98,02% | 97,51% | 97,79% | 97,60% | 97,51% | 93,94% | 95,28% |
| | Protein | 100% | 100% | 100% | 100% | - | 99,92% | 99,33% | 99,08% | 98,49% | 98,74% | 98,49% | 98,49% | 96,22% | 97,57% |
| Benin | DNA | 99,94% | 99,94% | 99,94% | 99,94% | 99,94% | - | 99,19% | 97,96% | 97,46% | 97,74% | 97,54% | 97,46% | 93,88% | 95,22% |
| | Protein | 99,92% | 99,92% | 99,92% | 99,92% | 99,92% | - | 99,24% | 98,99% | 98,41% | 98,66% | 98,41% | 98,41% | 96,14% | 97,48% |
| Mkuzi | DNA | 99,25% | 99,25% | 99,25% | 99,25% | 99,25% | 99,19% | - | 98,27% | 97,71% | 97,99% | 97,74% | 97,71% | 94,05% | 95,53% |
| | Protein | 99,33% | 99,33% | 99,33% | 99,33% | 99,33% | 99,24% | - | 99,16% | 98,57% | 98,83% | 98,66% | 98,57% | 96,22% | 97,48% |
| Georgia | DNA | 98,02% | 98,02% | 98,02% | 98,02% | 98,02% | 97,96% | 98,27% | - | 97,46% | 97,76% | 97,68% | 97,60% | 93,57% | 94,75% |
| | Protein | 99,08% | 99,08% | 99,08% | 99,08% | 99,08% | 98,99% | 99,16% | - | 98,99% | 99,50% | 98,83% | 99,08% | 96,22% | 97,48% |
| Tengani | DNA | 97,51% | 97,51% | 97,51% | 97,51% | 97,51% | 97,46% | 97,71% | 97,46% | - | 98,18% | 98,07% | 97,90% | 93,52% | 94,78% |
| | Protein | 98,49% | 98,49% | 98,49% | 98,49% | 98,49% | 98,41% | 98,57% | 98,99% | - | 99,33% | 98,91% | 99,24% | 96,22% | 97,65% |
| Pretoris | DNA | 97,79% | 97,79% | 97,79% | 97,79% | 97,79% | 97,74% | 97,99% | 97,76% | 98,18% | - | 98,85% | 98,60% | 93,60% | 95,03% |
| | Protein | 98,74% | 98,74% | 98,74% | 98,74% | 98,74% | 98,66% | 98,83% | 99,50% | 99,33% | - | 99,33% | 99,58% | 96,31% | 97,65% |
| Warm | DNA | 97,60% | 97,60% | 97,60% | 97,60% | 97,60% | 97,54% | 97,74% | 97,68% | 98,07% | 98,85% | - | 98,80% | 93,71% | 95% |
| | Protein | 98,49% | 98,49% | 98,49% | 98,49% | 98,49% | 98,41% | 98,66% | 98,83% | 98,91% | 99,33% | - | 99,08% | 96,56% | 97,57% |
| Warthog | DNA | 97,51% | 97,51% | 97,51% | 97,51% | 97,51% | 97,46% | 97,71% | 97,60% | 97,90% | 98,60% | 98,80% | - | 93,43% | 94,86% |
| | Protein | 98,49% | 98,49% | 98,49% | 98,49% | 98,49% | 98,41% | 98,57% | 99,08% | 99,24% | 99,58% | 99,08% | - | 96,06% | 97,40% |
| Kenya | DNA | 93,94% | 93,94% | 93,94% | 93,94% | 93,94% | 93,88% | 94,05% | 93,57% | 93,52% | 93,60% | 93,71% | 93,43% | - | 95,08% |
| | Protein | 96,22% | 96,22% | 96,22% | 96,22% | 96,22% | 96,14% | 96,22% | 96,22% | 96,22% | 96,31% | 96,56% | 96,06% | - | 97,23% |
| Malawi | DNA | 95,28% | 95,28% | 95,28% | 95,28% | 95,28% | 95,22% | 95,53% | 94,75% | 94,78% | 95,03% | 95% | 94,86% | 95,08% | - |
| | Protein | 97,57% | 97,57% | 97,57% | 97,57% | 97,57% | 97,48% | 97,48% | 97,48% | 97,65% | 97,65% | 97,57% | 97,40% | 97,23% | - |

Names of isolates were abbreviated as follows: OUR – OURT88/3 ; Benin – Benin07/1; Mkuzi – Mkuzi1979; Georgia – Georgia2007/1; Tengani – Tengani62; Pretoris – Pretorisuskop96/4; Warm – Warmbaths; Kenya – Kenya1950; Malawi – Malawi Lil/20/1

### 2.3.3 Phylogenetic analysis of viral type II topoisomerases

We also addressed the resemblance of P1192R to other known viral type II topoisomerases. Using the P1192R protein sequence as bait, we performed a protein-protein BLAST (BlastP) search to find similar enzymes in other viruses. We did not obtain many hits, and most are from closely related viruses. Those which are known to encode type II topoisomerases are all large DNA viruses, as is the case of ASFV, infecting a wide range of hosts, from prokaryotic to eukaryotic organisms, although none of them infects mammals. In the phylogenetic tree resulting from the comparison of these sequences (Figure 14) P1192R from L60 clusters with the topoII from phage T4 (composed by three different subunits but, for the purpose of this study, we concatenated the three sequences in order to yield a comparable topoII sequence), even though their difference is high as denoted by the branch lengths. The topoII from phage T4 is actually the sequence to which P1192R is less identical, and sequences to which it has higher identity are from algae-infecting viruses (see Figure 14).

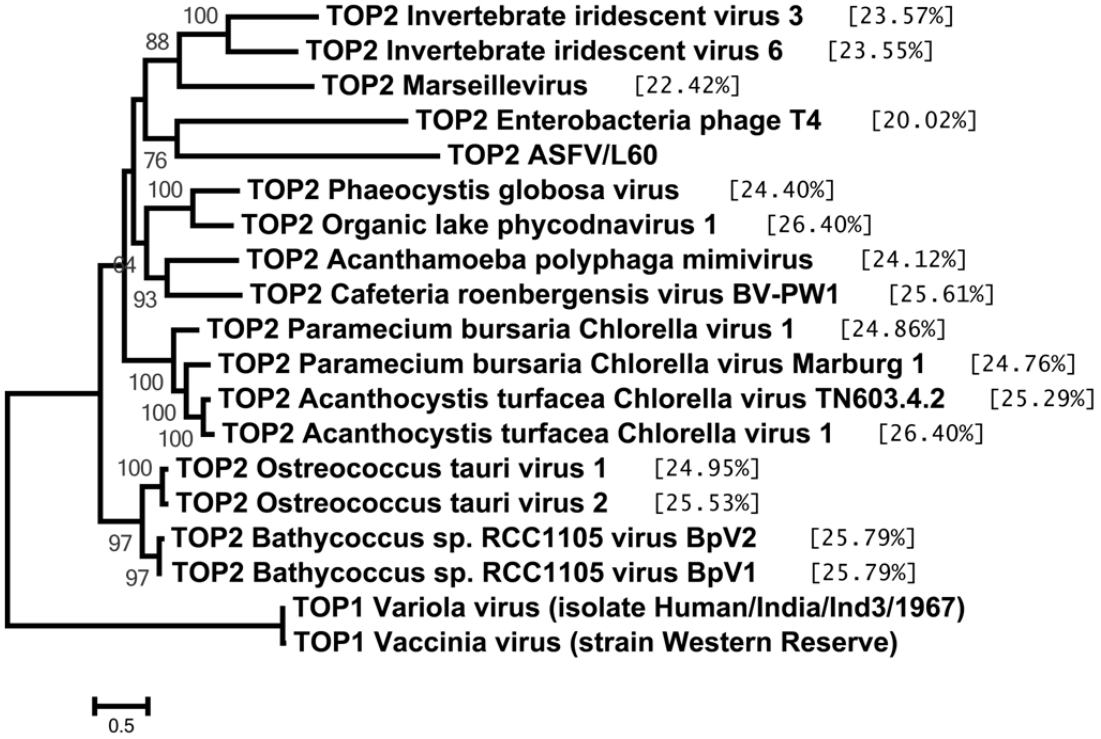


**Figure 14.** Maximum-likelihood phylogenetic analysis of type II topoisomerases from large DNA viruses.

Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of viral type II topoisomerases. Sequences for variola and vaccinia viruses are from type I topoisomerases and work as an outgroup. Values indicated between square brackets refer to percentage identity of the respective sequence to the ASFV/L60 sequence. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

### 2.3.4 Phylogenetic distribution of large DNA viruses sequences

One striking characteristic of large DNA viruses is that, in general, they rely very little on the cellular replication or transcription systems since they code for many of these essential components themselves. We wanted to know if among the NCLDVs other proteins would follow the same phylogenetic distribution that was observed for type II topoisomerases, to see if a common ancestry for the core of ASFV's machinery could be found. For this analysis we searched for possible essential proteins that would be (putatively) coded by all, or at least the majority, of the viral families of the NCLDVs. Given our criteria, we ended up analyzing viral homologs for PCNA, primase, DNA polymerase and RNA polymerase (resumed in Table 5).

The phylogenetic analysis of RNA polymerases (Figure 15) showed a distribution of the sequences in accordance with the family in which the respective viruses are classified. In this analysis, the RNApol from ASFV branches between sequences from families Poxviridae and Mimiviridae. Curiously, for the *Phaeocystis globosa* virus, from family Phycodnaviridae, two main RNA polymerase subunits were found and both were used in this analysis. While one of the sequences grouped with the remaining Phycodnaviridae, the other sequence branched close to sequences from viruses of family Marseilleviridae.

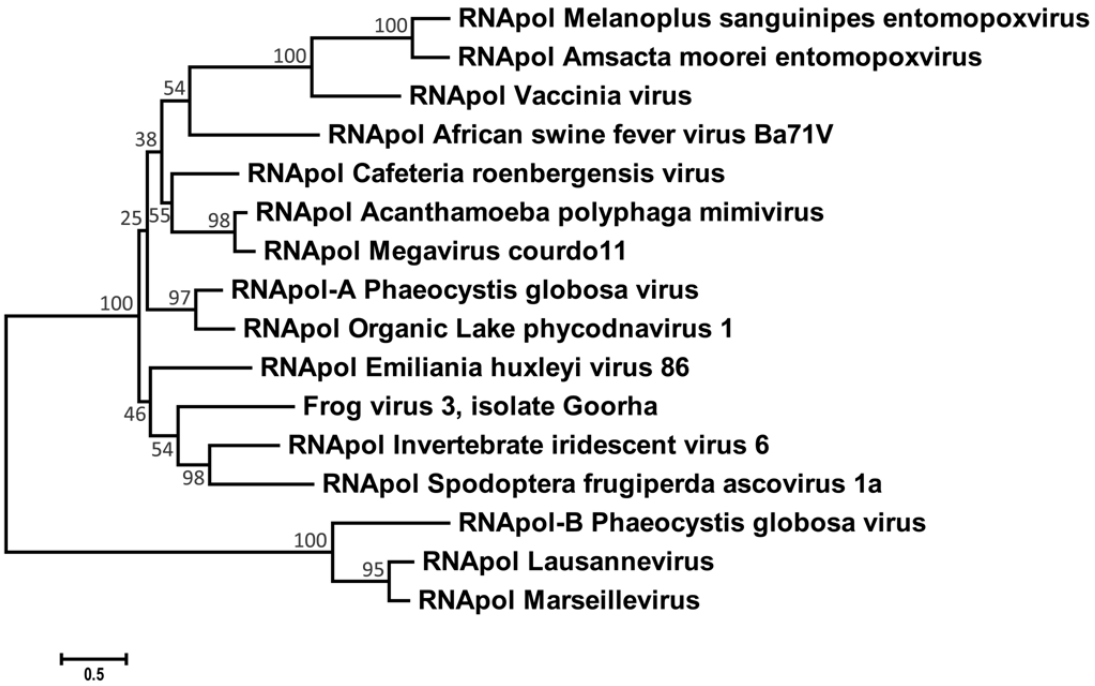


**Figure 15.** Maximum-likelihood phylogenetic analysis of RNA polymerase from large DNA viruses.

Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of viral RNA polymerases. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

**Table 5.** Components of replicative or transcriptional machineries present in the genome of large DNA viruses.

| | | Topoisomerase II | | PCNA | | DNA polymerase | | Primase | | RNA polymerase | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Family** | **Genus** | **viral ORF** | **Accession no.** | **viral ORF** | **Accession no.** | **viral ORF** | **Accession no.** | **viral ORF** | **Accession no.** | **viral ORF** | **Accession no.** |
| Ascoviridae | **Ascovirus** (*Spodoptera frugiperda* ascovirus 1a) | - | - | - | - | ORF1 | Q779J8 | ORF099 | YP_762453 | ORF008 | YP_762363 |
| Asfarviridae | **Asfivirus** (isolates L60 / Ba71V) | P1192R | KM261522 | E301R | NP_042820 | G1211R | NP_042783 | C962R | NP_042765 | NP1450L | NP_042792 |
| Iridoviridae | **Iridovirus** (Invertebrate iridescent virus 6) | IIV6-045L | Q9QSK1 | IIV6-436R | Q91F89 | IIV6-037L | Q9QSK2 | IIV6-184R | O55768 | IIV6-176R | O55766 |
| | **Lymphocystivirus** (Lymphocystis disease virus 1) | - | - | ORF45 | NP_078615 | - | - | - | - | - | - |
| | **Ranaviruses** (*Rana grylio* virus / Frog virus 3) | - | - | PCNA | ABR12624 | FV3-060R | Q6GZR5 | FV3-022R | Q6GZV3 | FV3-008R | Q6GZW7 |
| Marseilleviridae | **Lausannevirus** | LAU_0056 | YP_004347024 | LAU_0210 | YP_004347173 | LAU_0345 | YP_004347308 | LAU_0164 | YP_004347127 | LAU_0029 | YP_004346997 |
| | **Marseillevirus** | MAR_ORF173 | YP_003406916 | MAR_ORF212 | YP_003406953 | - | - | MAR_ORF152 | YP_003406895 | MAR_ORF023 | YP_003406778 |
| Mimiviridae | - (*Cafeteria roenbergensis* virus BV-PW1) | crov325 | YP_003969958 | crov219 | YP_003969851 | crov497 | YP_003970130 | crov494 | YP_003970127 | crov368 | YP_003970001 |
| | **Megavirus** (Megavirus courdo11) | CE11_00419 | AFX92446 | CE11_00401 | AFX92428 | CE11_00628 | AFX92654 | CE11_00310 | AFX92340 | CE11_00389 | AFX92416 |
| | **Mimivirus** (APMV) | MIMI_R480 | AEJ34722 | MIMI_R493 | AEJ34735 | MIMI_R322 | AEJ34561 | MIMI_L207/L206 | Q5UQ22.2 | MIMI_R501 | Q7T6X5 |
| Myoviridae | **T4-like viruses** (Enterobacteria phage T4) | gp39<br>gp52<br>gp60 | NP_049621<br>NP_049875<br>NP_049618 | gp45 | P04525 | gp43 | P04415 | gp61 | P04520 | - | - |
| Phycodnaviridae | **Chlorovirus** (*Paramecium bursaria Chlorella* virus 1) | A583L | Q5XLS4 | A193L<br>A574L | Q84513<br>O41056 | A185R | P30321 | A456L | Q98507 | - | - |
| | **Coccolithovirus** (*Emiliania huxleyi* virus 86) | EhV444 | YP_294202 | EhV020<br>EhV440 | YP_293774<br>YP_294198 | EhV030 | YP_293784 | - | - | EhV064 | YP_293818 |
| | **Phaeovirus** (*Ectocarpus siliculosus* virus 1) | - | - | ORF132 | NP_077617 | ORF93 | NP_077578 | ORF109 | NP_077594 | - | - |
| | **Phycodnavirus** (Organic Lake phycodnavirus 1) | 162322430 | ADX05889 | 162290251 | ADX06076 | 162285264 | ADX06143 | 162313527 | ADX06014 | 162313557 | ADX06029 |
| | **Prasinoviruses** (*Ostreococcus lucimarinus* virus OlV1) | OlV1_223c | YP_004061855 | OlV1_111 | YP_004061744 | OlV1_219 | YP_004061851 | OlV1_165c<br>OlV1_081 | YP_004061797<br>YP_004061714 | - | - |
| | **Prymnesiovirus** (*Phaeocystis globosa* virus) | PGCG_00051 | YP_008052370 | PGCG_00092 | YP_008052411 | PGCG_00248 | YP_008052566 | PGCG_00200 | YP_008052518 | PGCG_00263<br>PGCG_00228 | YP_008052581<br>YP_008052546 |
| Poxviridae | **Entomopoxviruses** (*Amsacta moorei* entomopoxvirus / *Melanoplus sanguinipes* entomopoxvirus) | - | - | - | - | AMV050<br>MSV036 | AAG02756<br>NP_048107 | AMV087<br>MSV089 | AAG02793<br>NP_048160 | AMV221<br>MSV043 | AAG02927<br>NP_048113 |
| | **Orthopoxvirus** (Vaccinia virus, strain Copenhagen) | - | - | VLTF1 | P68612 | E9L | P20509 | D5R | P21010 | J6R | P20504 |
| - | **Dinodnavirus** (*Heterocapsa circularisquama* DNA virus) | - | - | - | - | PolB | AB505427 | - | - | - | - |

In the phylogenetic analysis of DNA polymerases (Figure 16), ASFV's DNApol was found to be more closely related with its homolog from the recently described *Heterocapsa circularisquama* DNA virus, in accordance to what has been published (Ogata *et al.*, 2009). Unfortunately, very few sequences from this organism can be found in the databases, hindering us from determining if this similarity extends to other proteins coded by this virus. Still, both sequences branch before all of the other sequences, except the one from phage T4. Nevertheless, the ASFV DNApol sequence branches before the sequences of the Poxviridae, in accordance to what was seen in the RNApols analysis. In addition, although the sequences in this analysis are generally grouped in accordance to the viral family, as seen for the RNApols analysis, sequences from the Phycodnaviridae branch separately, with the sequences from Organic lake phycodnavirus 1 and from the *Phaeocystis globosa* virus grouping closer to the sequences from the Mimiviridae than to those of other Phycodnaviridae. The same was

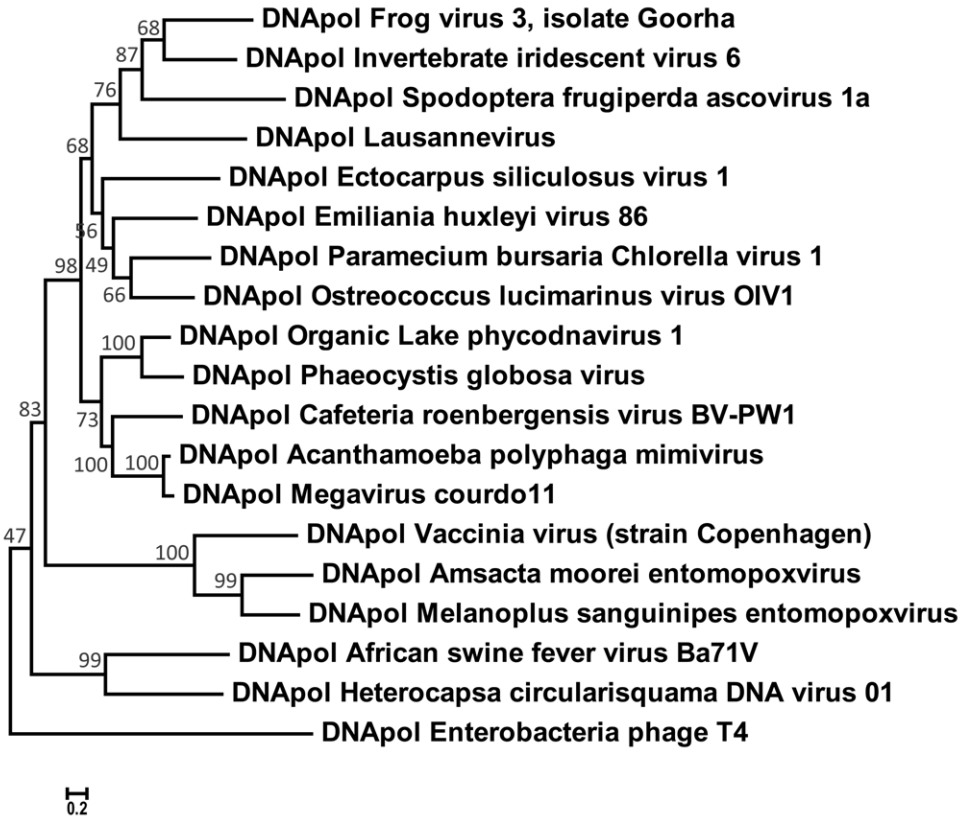


**Figure 16.** Maximum-likelihood phylogenetic analysis of DNA polymerase from large DNA viruses.

Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of viral DNA polymerases. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

observed in the phylogenetic analyses of viral primases (Figure 17) and viral PCNA homologs (Figure 18). In these analyses, while the ASFV primase sequence does not branch far from Poxviridae sequences, the ASFV PCNA homolog branches between sequences belonging to the Iridoviridae.

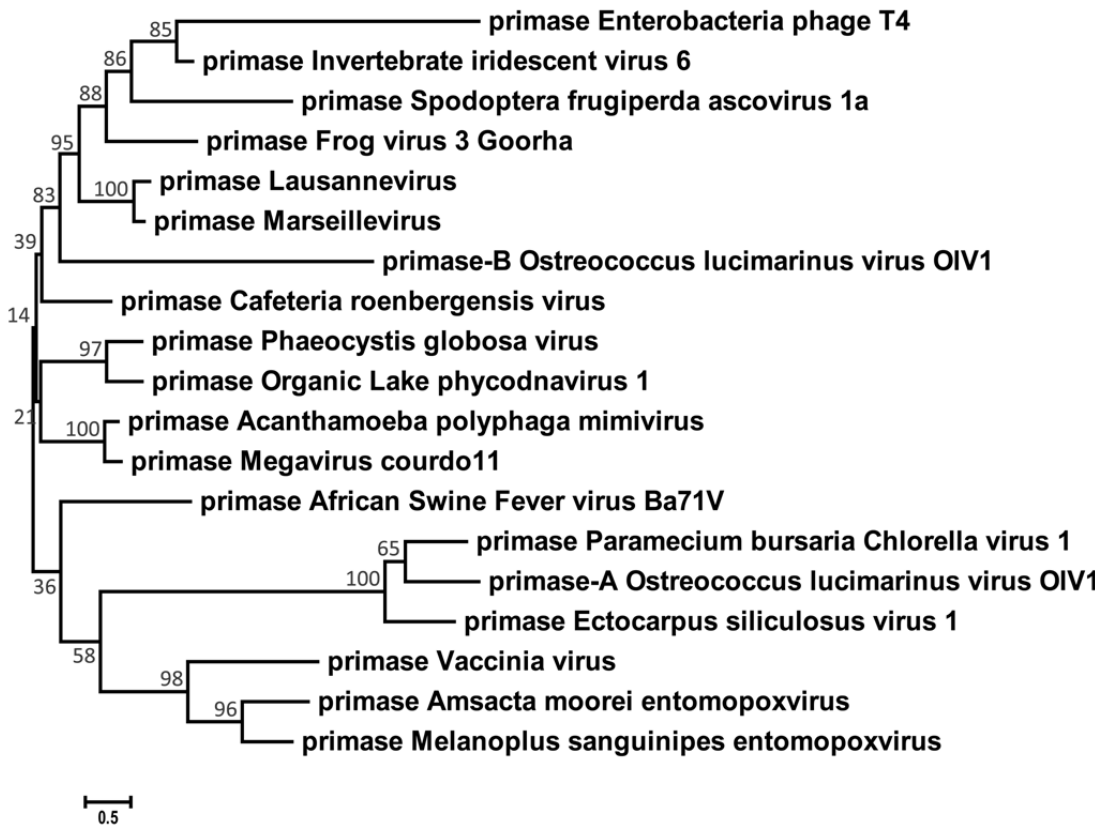


**Figure 17.** Maximum-likelihood phylogenetic analysis of primase from large DNA viruses. Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of viral primases. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

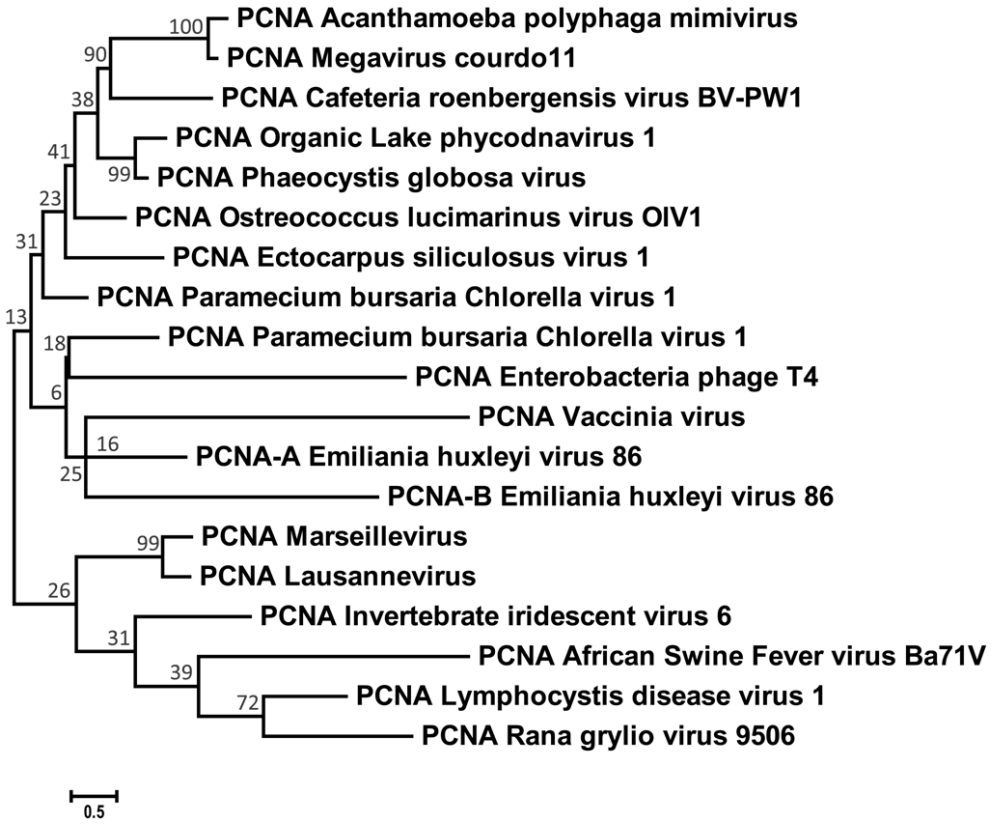


**Figure 18.** Maximum-likelihood phylogenetic analysis of PCNA homologs from large DNA viruses.

Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of viral homologs of PCNA. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

### 2.3.5 Comprehensive phylogenetic analysis of type II topoisomerases

Finally, we wanted to know how P1192R would fit in a broader picture of DNA topoisomerases. For this, we added to our comparison sequences from well-known and less-known topoII, as well as type I topoisomerases sequences which we expected to act as an outgroup, and we performed a new phylogenetic analysis. In the phylogenetic tree obtained (Figure 19) there is a clear separation between vertebrates and the remaining eukaryotes and, curiously, there's also a closer proximity of viral type II topoisomerases with those from parasitic organisms of family Trypanosomatidae. L60 P1192R branches in this latter cluster, but with considerable branch length, and it's somewhat bridging prokaryotic and eukaryotic topoII, indicating that it is highly divergent from known type II topoisomerases.

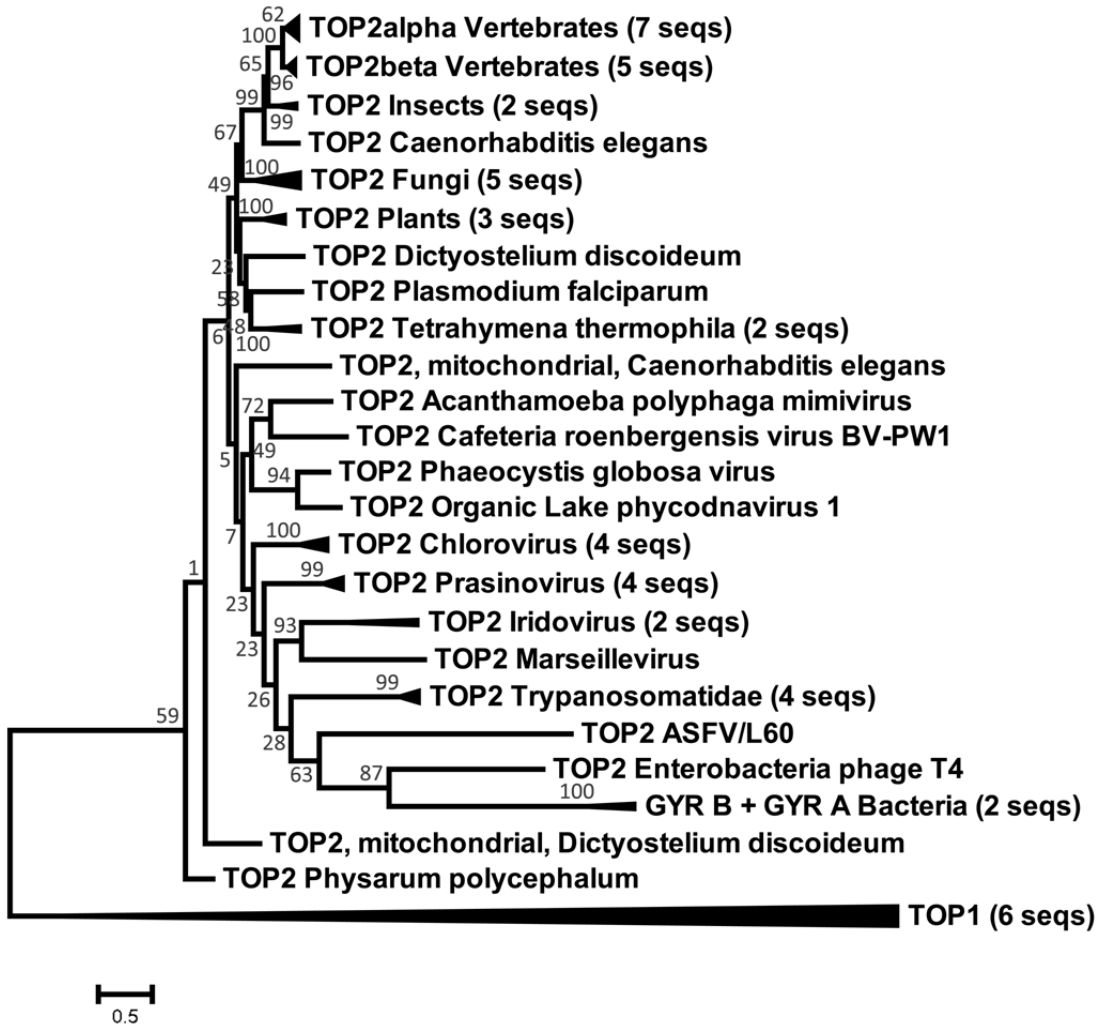


**Figure 19.** Maximum-likelihood phylogenetic analysis of type II topoisomerases from several domains of the tree of life.

Maximum-likelihood phylogenetic tree was constructed from a multiple amino acid sequence alignment of type II topoisomerases. Type I topoisomerases sequences work as an outgroup. The values between brackets denote the number of sequences contained in the collapsed branches. For details on these collapsed sequences, see Table 2. Bootstrap values are indicated in grey and the scale bar represents an estimated number of amino acid substitutions per site.

## 2.4 Discussion

We analyzed the putative type II topoisomerase P1192R from ASFV isolate Lisbon60 and found that several conserved motifs, present in prokaryotic and especially in eukaryotic type II topoisomerases, are present in P1192R, but the less-conserved C-terminal domain is absent (Figure 11). The importance of the C-terminal domain in enzymatic regulation has been previously reported, with sites and motifs present in this domain being responsible for regulating protein localization (Caron, Watt, & Wang, 1994; Jensen *et al.*, 1996; Mirski, Gerlach, & Cole, 1999; Mirski, Bielawski, & Cole, 2003), DNA-binding (Gilroy & Austin, 2011) and activity (Meczes, Gilroy, West, & Austin, 2008). Still, the C-terminal domain is also absent from some known type II topoisomerases, as in the case of topoII from *Chlorella* viruses, and these have been shown to have high DNA cleavage activity (Fortune *et al.*, 2001; Dickey, Choi, Van Etten, & Osheroff, 2005). Therefore, although it may be important, the C-terminal domain is not a requirement for enzymatic activity. In fact, the known (even if putative) type II topoisomerases that do not contain a C-terminal domain are virally encoded, by NCLDV, which suggests that the regulation occurring at the level of this domain may be of importance for when there are reduced copies of the genetic material, as in the case of prokaryotic or eukaryotic cells. In the cases of viruses, in which there are many copies of the genome waiting to be processed by the DNA topoisomerase II, these are probably required to have high levels of activity (as seen for the *Chlorella* viruses topoII) and the presence of the C-terminal domain could somehow diminish this activity by making the enzymes vulnerable to regulation like in their cellular counterparts. Nevertheless, the predicted tertiary and quaternary structures of pP1192R suggest that it structurally resembles known functional type II DNA topoisomerases, strengthening the hypothesis that ASFV indeed codes for a functional topoII. Interestingly, a pairwise alignment of the amino acid sequence of L60 pP1192R and of the yeast Top2p (see Figure 20), used as template by the RaptorX web-server to predict the structure of pP1192R, aligns pP1192R up to residue 1186 of Top2p (just before the start of its C-terminal region), and the model for structure prediction was the structure of a Top2p truncated at residue 1177 (therefore, also excluding the C-terminal region). Excluding the Top2p residues from 1187 onwards, that same pairwise alignment revealed that both sequences are only 25.6% identical (335 identical amino acid residues), with a similarity of 42.2% (551 amino acid residues). Thus, even though the sequences are very different, the domains and motifs essential for activity are highly conserved and the properties of the remaining amino acids allow for similar structural properties.

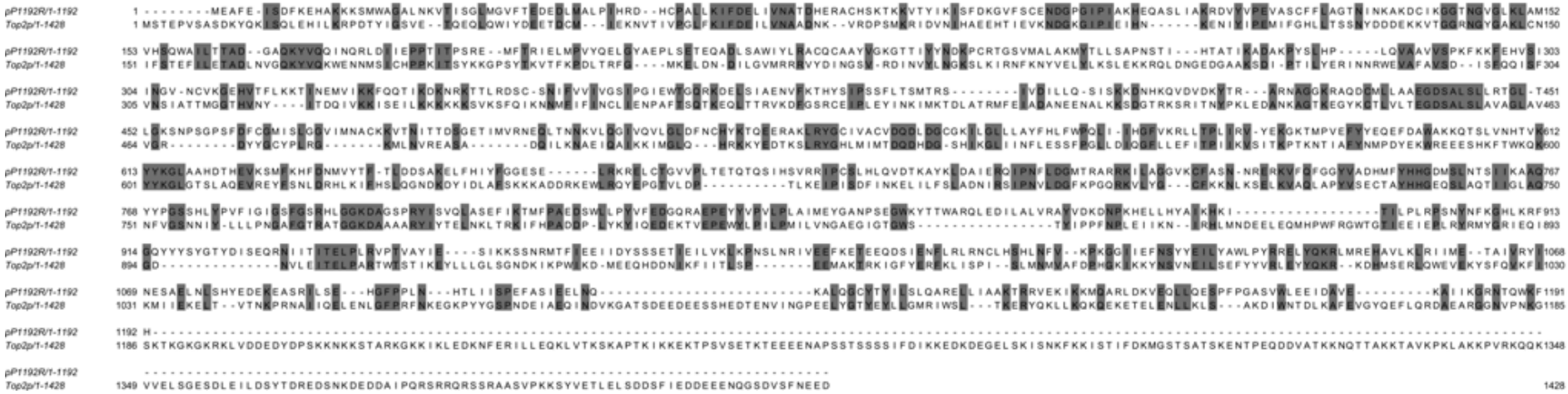

**Figure 20.** Pairwise alignment of pP1192R and Top2p.
Amino acid sequences of ASFV's pP1192R and *S. cerevisiae*'s Top2p were aligned using needle (available at http://www.ebi.ac.uk/Tools/psa/emboss_needle/). Shown in grey are identical residues between both sequences.

Our phylogenetic analysis of P1192R showed it to be highly conserved among the fourteen ASFV isolates studied (Figure 13), but quite distinct from other known type II topoisomerases (Figures 14 and 19). This implies that this protein either arose independently by action of convergent evolution or has evolved very rapidly, obscuring the determination of an eventual last common ancestor. If the selective pressure to which the protein (and thus its DNA sequence) is subjected is somewhat relaxed, it may allow a considerable level of variation in its sequence, when compared to other known type II topoisomerases, but still retaining the necessary motifs, organization and structural conformation required for function, as indicated by its predicted structure. As our phylogenetic analysis showed, this distinctiveness is not exclusive of P1192R but extends to other ASFV proteins that participate, or are predicted to do so, in viral replication and/or transcription.

Curiously, in the analysis of viral DNA polymerases, the ASFV sequence branches with the orthologous sequence from *Heterocapsa circularisquama* DNA virus (HcDNAV). *H. circularisquama* is a dinoflagellate, notable for the recurrent formation of toxic efflorescences, or blooms, with particular impact on the mortality of bivalves in Japan (Matsuyama, 1996). HcDNAV was described in 2001 (Tarutani, 2001) as being a large (≈ 356 kbp) cytoplasmic double-stranded DNA virus with a lytic infectious cycle. From their hosts and genome size it could seem that HcDNAV and ASFV would have little in common, and is thus remarkable to find such similarities at the protein level. It would be interesting to see if these similarities extend to other proteins of HcDNAV, perhaps hinting to a common origin for these two viruses. Conversely, in this phylogenetic analysis of DNA polymerases the sequence from phage T4 does not branch far from the ASFV sequence and it is possible to speculate that, in the absence of the HcDNAV sequence, the ASFV and phage T4 DNApol sequences could group together, as seen for their type II topoisomerases. However, this is not observed for the other proteins analyzed, since both primase and PCNA from these viruses are located in different branches of the respective phylogenetic trees (Figures 17 and 18, respectively). Since the ASFV and phage T4 proteins tend to have very low identity percentages, their pairing could simply be a result of long branch attraction. This is an error in phylogenetic analysis by which unrelated entities with high evolutionary rates are wrongly clustered together due to a biased-analysis of the evolutionary change (Brinkmann & Philippe, 2007). It arises when the molecular change accumulated in those entities is sufficient to make them appear similar, even though they are not related by descent. This usually translates into the formation of long branches within a phylogenic tree. Finally, despite usually branching before sequences from the Poxviridae, the ASFV sequences are usually isolated in the phylogenetic trees, with no frequent pairing, contrary to what is observed for the Mimivirus

and the Megavirus, the Marseillevirus and Lausannevirus, the Poxviridae or the majority of the Phycodnaviridae sequences. Therefore, despite our best efforts, we could not find a consensual common ancestor for ASFV proteins involved in essential mechanisms of the viral cycle.

## 2.5 References


Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10.

Baylis, S. A., Dixon, L. K., Vydelingum, S., & Smith, G. L. (1992). African swine fever virus encodes a gene with extensive homology to type II DNA topoisomerases. *Journal of Molecular Biology*, *228*(3), 1003–1010.

Berger, J. M., Gamblin, S. J., Harrison, S. C., & Wang, J. C. (1996). Structure and mechanism of DNA topoisomerase II. *Nature*.

Brinkmann, H., & Philippe, H. (2007). The Diversity Of Eukaryotes And The Root Of The Eukaryotic Tree. In *Eukaryotic Membranes and Cytoskeleton* (pp. 20–37).

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, *25*(15), 1972–3.

Caron, P. R., Watt, P., & Wang, J. C. (1994). The C-terminal domain of Saccharomyces cerevisiae DNA topoisomerase II. *Molecular and Cellular Biology*, *14*(5), 3197–207.

Chenna, R. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, *31*(13), 3497–3500.

Classen, S., Olland, S., & Berger, J. M. (2003). Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 10629–34.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)*, *27*(8), 1164–5.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

De Villiers, E. P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R., & Bishop, R. P. (2010). Phylogenomic analysis of 11 complete African swine fever virus genome sequences. *Virology*, *400*(1), 128–36.

Dickey, J. S., Choi, T.-J., Van Etten, J. L., & Osheroff, N. (2005). Chlorella virus Marburg topoisomerase II: high DNA cleavage activity as a characteristic of Chlorella virus type II enzymes. *Biochemistry*, *44*(10), 3899–908.

Dong, K. C., & Berger, J. M. (2007). Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature*, *450*(7173), 1201–1205.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–7.

Fass, D., Bogden, C. E., & Berger, J. M. (1999). Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands. *Nature Structural Biology*, *6*(4), 322-326.

Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, *25*(5), 471–492.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.

Feng, D.-F., & Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisitetto correct phylogenetic trees. *Journal of Molecular Evolution*, *25*(4), 351–360.

Fitch, W. (1971). Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology*, *20*(4), 406–416.

Fortune, J. M., Lavrukhin, O. V, Gurnon, J. R., Van Etten, J. L., Lloyd, R. S., & Osheroff, N. (2001). Topoisomerase II from Chlorella virus PBCV-1 has an exceptionally high DNA cleavage activity. *The Journal of Biological Chemistry*, *276*(26), 24401–8.

Gallardo, C., Mwaengo, D., Macharia, J., Arias, M., Taracha, E., Soler, A., Okoth, E., Martín, E., Kasiti, J., & Bishop, R. (2009). Enhanced discrimination of African swine fever virus isolates through nucleotide sequencing of the p54, p72, and pB602L (CVR) genes. *Virus Genes*, 38(1), 85–95.

García-Beato, R., Freije, J. M., López-Otín, C., Blasco, R., Viñuela, E., & Salas, M. L. (1992). A gene homologous to topoisomerase II in African swine fever virus. *Virology*, *188*(2), 938–47.

Gilroy, K. L., & Austin, C. a. (2011). The impact of the C-terminal domain on the interaction of human DNA topoisomerase II α and β with DNA. *PloS One*, *6*(2), e14693.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, *38*(Web Server issue), W695–9.

Guindon, S., Dufayard, J.-F. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321.

Hillis, D. M., & Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology , 42* (2 ), 182–192.

Jensen, S., Andersen, a H., Kjeldsen, E., Biersack, H., Olsen, E. H., Andersen, T. B., Westergaard, O., & Jakobsen, B. K. (1996). Analysis of functional domain organization in DNA topoisomerase II from humans and Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *16*(7), 3866–77.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Research*, *35*(Web Server issue), W429–32.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, *7*(8), 1511–1522.

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066.

La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., & Brunak, S. (2004). Analysis and prediction of leucine-rich nuclear export signals. *Protein Engineering, Design & Selection : PEDS*, *17*(6), 527–36.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, *23*(21), 2947–8.

Mandel, M., & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology*, *53*(1), 159–162.

Matsuyama, Y. (1996). Biological and environmental aspects of noxious dinoflagellate red tides by Heterocapsa circularisquama in the West Japan. *In Harmful and Toxic Algal Blooms.*, 247–250.

Meczes, E. L., Gilroy, K. L., West, K. L., & Austin, C. a. (2008). The impact of the human DNA topoisomerase II C-terminal domain on activity. *PloS One*, *3*(3), e1754.

Mirski, S. E. ., Bielawski, J. C., & Cole, S. P. . (2003). Identification of functional nuclear export sequences in human topoisomerase IIα and β. *Biochemical and Biophysical Research Communications*, *306*(4), 905–911.

Mirski, S. E. L., Gerlach, J. H., & Cole, S. P. C. (1999). Sequence Determinants of Nuclear Localization in the α and β Isoforms of Human Topoisomerase II. *Experimental Cell Research*, *339*, 329–339.

Morais Cabral, J. H., Jackson, a P., Smith, C. V, Shikotra, N., Maxwell, a, & Liddington, R. C. (1997). Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature*, *388*(6645), 903–906.

Moretti, S., Armougom, F., Wallace, I. M., Higgins, D. G., Jongeneel, C. V, & Notredame, C. (2007). The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Research*, *35*(Web Server issue), W645–8.

Nakai, K., & Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences*, *24*(1), 34-6.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–17.

Ogata, H., Toyoda, K., Tomaru, Y., Nakayama, N., Shirai, Y., Claverie, J.-M., & Nagasaki, K. (2009). Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. *Virology Journal*, *6*, 178.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research*, *33*(Web Server issue), W116–20.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(6), 276–277.

Rzhetsky, A., & Nei, M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution* , *10* (5 ), 1073–1095.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Schmidt, B. H., Osheroff, N., & Berger, J. M. (2012). Structure of a topoisomerase II-DNA-nucleotide complex reveals a new control mechanism for ATPase activity. *Nature Structural & Molecular Biology*, *19*(11), 1147–54.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, *28*, 1409–1438.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary

distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–9.

Tarutani, K. (2001). Isolation of a virus infecting the novel shellfish-killing dinoflagellate Heterocapsa circularisquama. *Aquat. Microb. Ecol.*, *23*, 103–111.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

Tretter, E. M., Schoeffler, A. J., Weisfield, S. R., & Berger, J. M. (2010). Crystal structure of the DNA gyrase GyrA N-terminal domain from Mycobacterium tuberculosis. *Proteins*, *78*(2), 492–495.

Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Molecular Biology and Evolution* , *14* (7 ), 717–724.

Yu, C.-S., Chen, Y.-C., Lu, C.-H., & Hwang, J.-K. (2006). Prediction of protein subcellular localization. *Proteins*, *64*(3), 643–51.

Yuan, Z., Mattick, J. S., & Teasdale, R. D. (2004). SVMtm: support vector machines to predict transmembrane segments. *Journal of Computational Chemistry*, *25*(5), 632–6.

# CHAPTER 3

# STUDIES ON THE EXPRESSION OF P1192R IN A CELLULAR CONTEXT

## 3.1 Introduction

Topoisomerases are essential enzymes required to resolve topological problems in DNA molecules. To fulfill their role they must co-localize with their substrate. This is facilitated in bacterial cells, due to the lack of membrane compartmentalization, in which regulation of these enzymes seems to occur on the basis of protein-protein interactions and of accumulation of topological structures on DNA molecules. Even so, bacterial topoisomerases still differ in their localization according to their specific cellular role and to the progression of the cell cycle (Espeli, Levine, Hassing, & Marians, 2003; Tadesse & Graumann, 2006). As bacterial type II topoisomerases are heterotetramers composed of different subunits (ParC/ParE for topoIV and GyrA/GyrB for DNA gyrase), when they are not assembled the localization of one subunit frequently contrasts with the localization of its respective pair (Espeli *et al.*, 2003). In eukaryotic cells, organelle compartmentalization complicates the regulation of topoisomerases localization, which is mostly accomplished through the presence of nuclear localization signals (NLSs). While topoI contains NLSs on its N-terminal region (Alsner, Svejstrup, Kjeldsen, Sørensen, & Westergaard, 1992; Mo, Wang, & Beck, 2000), topoIIs have NLSs on their C-terminal domain (Caron, Watt, & Wang, 1994; Jensen *et al.*, 1996; Adachi *et al.*, 1997; Mirski *et al.*, 1997; Mirski, Gerlach, & Cole, 1999; Mirski, Bielawski, & Cole, 2003), which is the most divergent domain in eukaryotic type II topoisomerases. However, regulation of topoIIs is not strictly dependent on these localization signals but is also cell cycle-related, since mammalian isoforms IIα and IIβ differ in their subcellular localization and expression pattern along the cell cycle (Austin & Marsh, 1998). Alteration of the cellular localization of type II topoisomerases is one way by which cancer cells gain resistance to topoII-targeting drugs like etoposide. This change in localization can occur either by active transport of the protein from the nucleus, via the CRM-1 transporter (Engel *et al.*, 2004), or by expression of an isoform truncated in the C-terminal region, thereby losing its respective NLSs (Wessel *et al.*, 1997).

Aside the purification and *in vitro* characterization of type II topoisomerases from the T-family of enterophages using infected *E. coli* cells (Liu, Liu, & Alberts, 1979; Huang, Wei, & Casjens, 1985), type II topoisomerases from *Chlorella* viruses have also been purified after heterologous expression in *S. cerevisiae* and characterized *in vitro* (Lavrukhin *et al.*, 2000; Dickey, Choi, Van Etten, & Osheroff, 2005). A type I topoisomerase from vaccinia virus has also been purified from virions (Shaffer & Traktman, 1987), while a topoI from mimivirus was expressed and purified from *E. coli* (Benarroch, Claverie, Raoult, & Shuman, 2006), and both had their activity characterized *in vitro*. Nevertheless, no studies have focused on the *in vivo* localization of these topoisomerases and on their exact role during infection.

Therefore, the work described in this chapter aims to characterize the protein coded by ORF P1192R in a cellular context, hoping to shed some light on its role during ASFV infection.

## 3.2 Materials and methods

### 3.2.1 Mammalian cell lines

Vero E6 (ECACC), COS-1 CRL-1650 (ATCC) and COS-7 CRL-1651 (ATCC) cell lines were grown at 37 ºC under a 5% $CO_2$ atmosphere saturated with water vapor, in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/l Glucose, 0.11 g/l Sodium Pyruvate and 2 mM L-gluthamine, supplemented with 10% (v/v) inactivated fetal bovine serum, penicillin/streptomycin (100 units/ml and 100 µg/ml, respectively) and nonessential amino acids (all from Gibco, Life Technologies). All cell lines were confirmed free of mycoplasma.

Pig macrophages were obtained as previously described (Portugal, Leitão, & Martins, 2009). Briefly, heparinized blood samples from crossbred Large White x Landrace pigs (6-month old) were obtained during bleeding at the abattoir. Samples were incubated at 37 ºC for 15 minutes with 10% (v/v) of a 5% (v/v) Dextran T500 (Sigma-Aldrich) solution in Hank's balanced saline solution (Gibco, Life Technologies). Supernatants were collected, diluted 1:1 in culture medium [RPMI 1640, supplemented as necessary to obtain final concentrations of 100 units/ml penicillin, 100 µg/ml streptomycin and 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (all from Gibco, Life Technologies)], and seeded (100 ml/flask) in T175 tissue culture flasks. Cultures were incubated as above for 48 hours and non-adherent cells were removed by extensive washing in pre-warmed phosphate-buffered saline (PBS, Gibco, Life Technologies).

### 3.2.2 Viruses and infections

ASFV isolates used were the highly virulent ASFV/L60 (Leitão *et al.*, 2001) (L60) and the Vero-adapted ASFV/Ba71V (Yáñez *et al.*, 1995) (Ba71V). L60 or Ba71V suspensions, purified as described by Leitão *et al.* (2000), were titrated by observation of cytopathic effect at end-point dilutions in macrophage cultures or Vero monolayers, respectively, and expressed as 50% tissue culture infectious dose ($TCID_{50}$).

Purified Ba71V or L60 were added to Vero/COS-7 cells or pig macrophages, respectively, at the indicated multiplicities of infection (MOI), and after 1 hour of adsorption at 37 °C the inoculum with the non-adsorbed virus was removed and cells were washed with serum-free medium and further incubated with complete culture medium up to the indicated time periods.

### 3.2.3 Cloning of P1192R from ASFV/L60

All procedures required for cloning were performed as stated on Chapter 2. Modifications made to the methods are indicated when necessary.

For expression in mammalian cells, P1192R was cloned in several vectors, as described below. In all of them, expression was driven by the human cytomegalovirus immediate early promoter. One of the vectors used was pIC113 (Cheeseman & Desai, 2005) in which P1192R was cloned in frame with the existing GFP ORF, allowing the construction of an N-terminal GFP-P1192R fusion. For cloning in this plasmid, P1192R was PCR-amplified (without the endogenous ATG) with primers including sites for *Bsp*EI and *Kpn*I and cloned in pIC113 using these enzymes. To obtain the mutant versions of P1192R, without the putative NLS or the putative nuclear export sequence (NES), site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies), according to the manufacturer's instructions, including the primer design. Creation of the NLS mutant implied the deletion of nucleotides 877 to 897, while for the NES mutant the deleted nucleotides were from 1327 to 1356. In both cases, the reading frame was maintained. Vectors pIC111 and pIC112 (Cheeseman & Desai, 2005) were also used for expression of P1192R in mammalian cells. In these vectors, ORF P1192R was cloned in frame with the existing GFP or RFP ORF, respectively, allowing the construction of a C-terminal P1192R-GFP or -RFP fusion. For cloning in these vectors, ORF P1192R was PCR-amplified (with the endogenous ATG but skipping the endogenous stop codon) with primers including sites for *Nhe*I and *Not*I and cloned using these enzymes.

The primers used in this chapter are listed below in Table 6. Sequences of restriction sites included in the primers are underlined. Nucleotides included upstream of those sequences were added to increase efficiency of the hydrolysis reactions. The primers were synthesized either by Life Technologies or by STAB VIDA. Sequencing was performed using the same strategy indicated on Chapter 2.

**Table 6.** Primers used in this chapter.

| Primer name | Sequence (5' → 3') | Purpose |
|---|---|---|
| P1192R noStop NotI Low | CCGCGGCCGCATGAAATTTCCACTGAGTA | cloning |
| P1192R 2nd Frag FW | CCGAATTCCCCAATTACAATTTCAAGGGCCATTTGAAGCGGTTTGGCCA | sequencing |
| P1192R iSeq1 Up | ACAAATCAACCAGCGCCT | sequencing |
| P1192R iSeq3 Up | CGACGAATTCCTTGCAGCCT | sequencing |
| P1192R iSeq4 Up | CAACAGACGATTAAAGATAAAAACC | sequencing |
| P1192R iSeq5 Up | GAGCATGGATTTCCCCCGCTG | sequencing |
| P1192R iSeq6 Low | CTATGGCAAGCTCTTTCATGATCC | sequencing |
| RtPr2 P1192R Up | AGCGAGCAAAGCTGAGATACG | sequencing |
| T7_RV_JC | GCTAGTTATTGCTCAGCGG | sequencing |
| T7_FW_JC | TAATACGACTCACTATAGGG | sequencing |
| P1192R BspEI noATG Up | CCTCCGGAGAAGCGTTTGAAATCAGCGATTTCTTTCAA | cloning |
| P1192R NheI Up JC | CCGCTAGCCATGGAAGCGTTTGAAATCAGCGATTTC | cloning |
| P1192R KpnI Low | CCGGTACCTTAATGAAATTTCCACTGAGTATTTCTTCC | cloning |
| del293-299_antisense | GATAATGGACACGTGGGACACGACCGCCGC | mutagenesis |
| del293-299_sense | GCGGCGGTCGTGTCCCACGTGTCCATTATC | mutagenesis |
| del443-452_antisense | GGTTGGACTTTCCTGCGCTATCCCCTTCCG | mutagenesis |
| del443-452_sense | CGGAAGGGGATAGCGCAGGAAAGTCCAACC | mutagenesis |

Sequences recognized by restriction enzymes are shown underlined – GCGGCCGC for *Not*I, GAATTC for *Eco*RI, TCCGGA for *Bsp*EI, GCTAGC for *Nhe*I and GGTACC for *Kpn*I. Some of the primers used for sequencing were also used for colony PCR.

### 3.2.4 Electroporation

In order to transfect cells by electroporation, sub-confluent monolayers were harvested by trypsinization, washed twice with cold PBS and resuspended in 2 ml of ice-cold electroporation buffer (120 mM KCl, 10 mM $KH_2PO_4$, 10 mM $K_2HPO_4$, 2 mM EGTA, 5 mM $MgCl_2$, 25 mM HEPES, 0.15 mM $CaCl_2$, pH 7.4). Per electroporation, 300 µl of the cell suspension was mixed with 30 µg of plasmid DNA, the mix was transferred to a 0.2 cm electroporation cuvette and incubated for 10 minutes on ice. Each cuvette was subjected to a pulse of 700 V and 50 µF capacitance (using a Gene Pulser II electroporator, Bio-Rad), followed by a brief period on ice until the addition of complete culture medium, to resuspend the cells. These were then plated on 24-well tissue culture plates, at the desired cell density.

### 3.2.5 Immunofluorescence

Monolayers of Vero, COS-1 or COS-7 cells grown in 24-well tissue culture plates containing coverslips, at 1 x $10^5$ viable cells per well, were inoculated with Ba71V, at a MOI of 2. Where indicated, the cells were incubated with 20 ng/ml of leptomycin B

(Sigma-Aldrich) in complete culture medium for 3 hours prior to fixation, or with 50 µg/ml of cytosine arabinoside (Sigma-Aldrich) in complete culture medium for 4 hours or 16 hours prior to fixation. Adherent macrophages were harvested after incubation with cold 0.8 mM EDTA in PBS, washed with PBS, and plated at 1.5 x $10^5$ cells per well, allowed to adhere for 3 hours, and washed again to remove non-adherent cells. They were then inoculated with L60, at a MOI of 2.

Cells were then washed twice with PBS and fixed with a 3.7% (w/v) paraformaldehyde solution (in PBS) for 10 minutes, with slight agitation. After two five-minutes washes with PBS, cells were permeabilized with PBS containing 0.1% (w/v) Triton X-100 for two minutes, washed again twice with PBS for five minutes and finally washed with PBS containing 0.1% (w/v) Tween 20. Cells were then blocked with PBS containing 3% (w/v) bovine serum albumin (BSA, Sigma-Aldrich) (blocking solution) for thirty minutes, after which they were incubated for one hour with the primary antibody(ies). Two washes with PBS and one with PBS containing 0.1% (v/v) Tween 20 followed, as did a one hour incubation with the secondary antibody(ies). After two washes with PBS, cells were incubated for five minutes with PBS containing 4 µg/ml of 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI, Sigma-Aldrich). Finally, coverslips were washed twice with PBS for 10 minutes before being mounted in DAPI-containing VECTASHIELD (Vector Laboratories), in which case the DAPI staining was omitted, or in Fluoromount-G (SouthernBiotech). All the incubations described above were performed at room temperature.

### 3.2.6 Topoisomerase II retention assay

Monolayers of Vero cells grown in 24-well tissue culture plates, at 1 x $10^5$ viable cells per well, were inoculated with Ba71V, at a MOI of 2. Additionally, incubation was performed without or with 100 µg/ml enrofloxacin (Sigma-Aldrich), added at 2 hours post-infection (hpi). Infection was allowed to proceed for 6 or 12 hours. Before fixation, cells were extracted on ice, with very gentle agitation, for 1 - 1.5 minutes with HPEM buffer (30 mM HEPES, 65 mM Pipes, 10 mM EGTA, 2 mM $MgCl_2$, pH 6.9) supplemented with 350 mM NaCl, 0.5% (w/v) Triton X-100, 1 mM phenylmethylsulphonyl fluoride (PMSF) and a commercially available mixture of protease inhibitors (Complete Mini EDTA-free, Roche Diagnostics). Finally, the cells were fixed in 3.7% formaldehyde in HPEM for 10 min at room temperature and processed as described in Section 3.2.5.

### 3.2.7 Immunoblotting

Monolayers of Vero cells grown in 24-well tissue culture plates, at 1 x $10^5$ viable cells per well, were inoculated with Ba71V, at a MOI of 10. At the desired times, the culture medium was collected, cells were washed twice with PBS 1x, collected by trypsinization and pooled with the respective culture medium. After centrifugation at 500 *g* for 5 minutes, 4 ºC, cells were washed once with PBS 1x and resuspended in lysis buffer [10 mM Tris-HCl pH 8.0, 5 mM EDTA, 1% (w/v) sodium dodecylsulphate (SDS)]. After a short-period of incubation on ice, cell suspensions were subjected to sonication and the cell lysates were analyzed by polyacrylamide gel electrophoresis (PAGE) in the presence of SDS (SDS-PAGE), using commercially available pre-cast polyacrylamide gels (Bio-Rad). Before being loaded in the gels, protein samples were mixed with 5x protein loading buffer [125 mM Tris-HCl, pH 6.8, 50% (v/v) glycerol, 10% (w/v) SDS, 0.25% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol] and heated at 100 ºC for 5 - 10 minutes, followed by centrifugation at 13000 rpm for 1 - 2 minutes in a microcentrifuge. The gels were run in Tris/Glycine/SDS buffer [0.025 M Tris-Base, 0.192 M glycine, 0.1% (w/v) SDS] at 90 V. The molecular weight marker used was NZYColour Protein Marker II (NZYTech).

Transfer of proteins from polyacrylamide gels to nitrocellulose membranes (0.22 µm Whatman, GE Life Sciences) was performed using the Mini Trans-Blot electrophoretic transfer cell system (Bio-Rad), according to the recommendations of the manufacturer. Transfer was done at 100 V for 1.5 hours in ice-cold transfer buffer [25 mM Tris base, 192 mM glycine, 20% (v/v) methanol, 0.1% (w/v) SDS] with slight agitation. The transfer buffer contained SDS to promote transfer of high molecular weight proteins. Success of the procedure was confirmed by staining membranes with Ponceau-S solution (Sigma-Aldrich).

Immunoblots were blocked for 2 - 3 hours with PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM $Na_2HPO_4$ * $2H_2O$, 1.8 mM $KH_2PO_4$) containing 5% non-fat milk, 5% yeast extract and 0.1% Tween 20, before being probed with primary and secondary antibodies or anti-sera. Finally, membranes were developed using chemiluminescent reagents SuperSignal West Pico or SuperSignal West Femto (ThermoScientific), according to the manufacturer`s instructions.

### 3.2.8 Antibodies and antisera

For immunofluorescence studies, the primary antibodies used were a mouse polyclonal anti-pP1192R serum, raised against recombinant full-length pP1192R expressed in *S. cerevisiae* (described in detail in Chapter 4), an in-house swine anti-ASFV whole-serum, an anti-vp32 antiserum kindly provided by Dr. Michael Parkhouse, Instituto Gulbenkian de Ciência, a commercial anti-vp72 antibody (M.11.PPA.I1BC11, Ingenasa), and a commercial

anti-NF-κB p65 rabbit polyclonal antibody (sc-7151, Santa Cruz Biotechnology), used at 1:1000, 1:500, 1:50, 1:50 and 1:100 dilutions, respectively. The conjugated secondary antibodies were the following: anti-pig-FITC IgG (ab6773, Abcam), anti-pig-TexasRed IgG (ab6775, Abcam) and anti-mouse-Cy3 IgG (A10521, Life Technologies), all at 1:500, and anti-rabbit-AlexaFluor 647 IgG (A-21244, Life Technologies), at 1:200. All dilutions of primary and secondary antibodies were performed in blocking solution.

For Western blot analyses, the primary antibodies used were the same mouse polyclonal anti-pP1192R serum and the in-house swine anti-ASFV whole-serum, both at 1:1000. The conjugated secondary antibodies were an anti-mouse-HRP conjugated (A2554, Sigma-Aldrich), and an anti-swine-HRP conjugated (114-035-003, Jackson Immunoresearch), at 1:50000 and 1:60000, respectively. All dilutions of primary and secondary antibodies were performed in blocking solution.

## 3.3 Results

### 3.3.1 Subcellular localization of GFP-pP1192R

To assess the localization of pP1192R, we constructed a GFP-P1192R fusion in order to analyze the distribution of the fusion protein by fluorescence microscopy in transfected cells. However, both cationic lipid transfection and electroporation of Vero cells were troublesome, associated with high cell mortality and with very low transfection efficiencies. To overcome this we started using COS-1 and COS-7 cells. These lines derive from CV-1 cells by transformation with modified SV40 virus and express the SV40 large T antigen (Gluzman, 1981) which confers them the capacity to replicate to high copy number plasmids containing the SV40 replication origin, as is the case of the ones we have been using for expression in mammalian cells lines. Also, these lines have been recently described as capable of sustaining infection by the Vero-adapted isolate ASFV/Ba71V (de León, Bustos, & Carrascosa, 2013).

A comparison of Vero, COS-1 and COS-7 transfected cells determined the latter as those with best transfection efficiencies, under the conditions tested. Therefore, these were chosen for subsequent studies. While COS-7 cells expressing GFP alone, used as controls, present a uniform staining both in the cytoplasm and in the nucleus, we observed that GFP-pP1192R is cytoplasmically distributed (Figure 21), with an apparent nuclear exclusion. This was also observed using P1192R-GFP or -RFP C-terminal fusions, but because the stability and/or expression levels of these recombinant proteins were much lower, we henceforth used only the N-terminal GFP fusion. Since the virus may need its type II topoisomerase for early gene expression, and given the fact that viral genomes are described as being present in the nucleus at early times of infection (Caeiro, Meireles, Ribeiro, & Costa, 1990; Brookes, Dixon, &

Parkhouse, 1996; Eulálio *et al.*, 2007; Ballester *et al.*, 2010), we wanted to clarify if pP1192R was indeed excluded from the nucleus. Given that our bioinformatic analysis of P1192R identified putative nuclear localization and nuclear export signals, we independently deleted

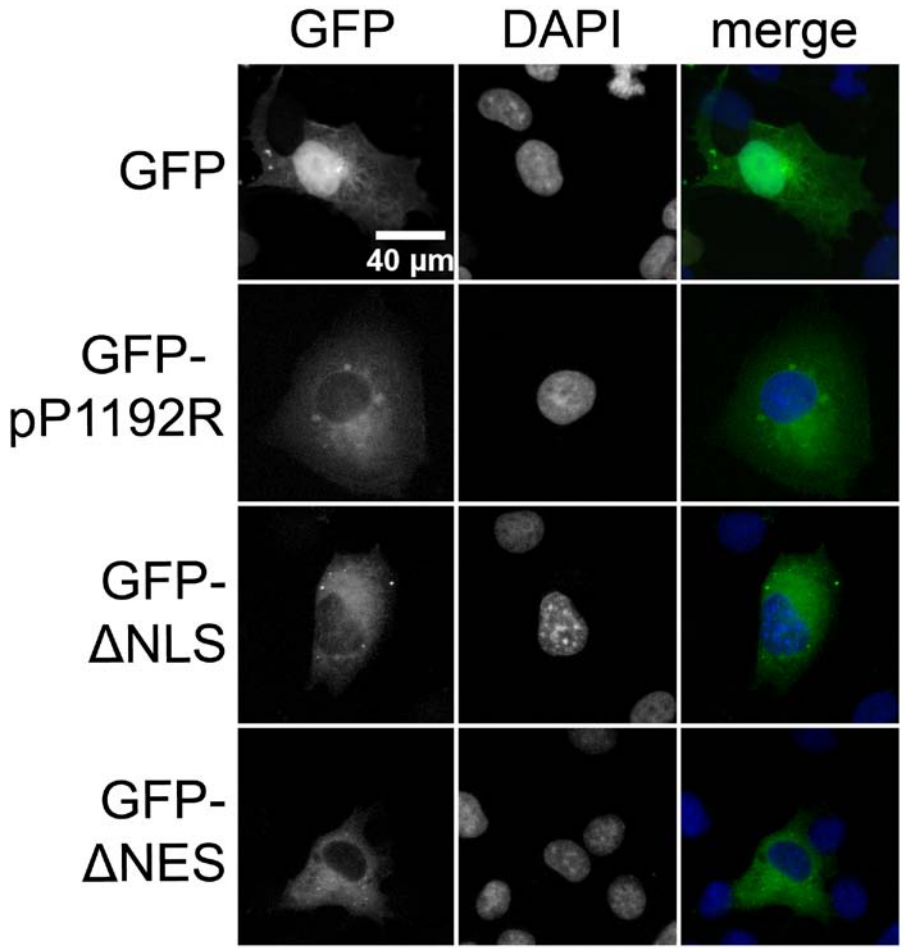


**Figure 21.** Evaluation of the cellular distribution of GFP-pP1192R.
COS-7 cells were transfected with GFP, GFP-P1192R or respective mutant forms in which the sequences for the putative nuclear localization signal or nuclear export signal were deleted (ΔNLS and ΔNES, respectively). In the merged panels, the GFP signal is shown in green and DNA DAPI staining in blue. Representative microscopy images of at least two independent experiments are shown.

these sequences in the GFP-P1192R fusion to test if the cellular distribution of the resulting protein would change. However, none of these deletions altered the localization of the fusion protein, which remained cytoplasmic (Figure 21).

In addition, we performed a 3-hour treatment of GFP-pP1192R transfected cells with leptomycin B, which specifically inhibits the CRM-1-dependent nuclear export pathway, but the treatment had no effect on the observed localization of GFP-pP1192R, even though it led to the nuclear accumulation of the p65 subunit of NF-κB, used as a control (Figure 22).

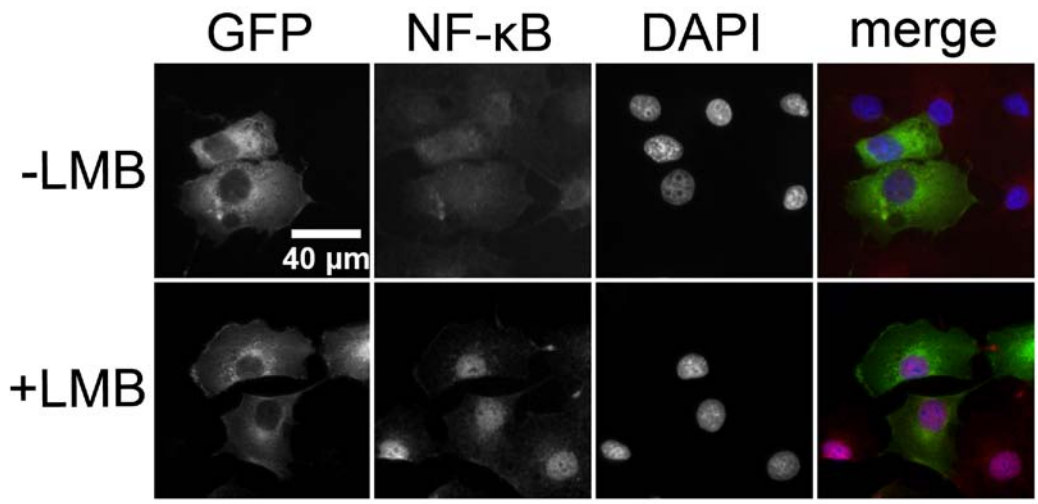


**Figure 22.** Effect of leptomycin B on the cellular distribution of GFP-pP1192R.
COS-7 cells were transfected with GFP-P1192R and, where indicated, treated with 20 ng/ml leptomycin B (LMB) for 3 hours. In the merged panels, the GFP signal is shown in green, the NF-κB anti-p65 signal in red and DNA DAPI staining in blue. Representative microscopy images of at least two independent experiments are shown.

The cellular localization of pP1192R could be influenced, or regulated, by other viral protein(s), or even by the presence of viral DNA. Thus, we decided to infect COS-7 transfected cells with Ba71V. At 4 hpi we didn't observe any change in the cytoplasmic distribution of GFP-pP1192R, even after incubation with leptomycin B (Figure 23). However, at 8hpi we observed a striking accumulation of GFP-pP1192R in the viral factories, together with a reduction of the GFP signal throughout the cytoplasm, implying that the recombinant protein is recruited during infection. This recruitment remained unaltered after leptomycin B incubation. Infection of COS-7 cells transfected with GFP alone had no effect on the localization of the fluorescent protein (not shown).

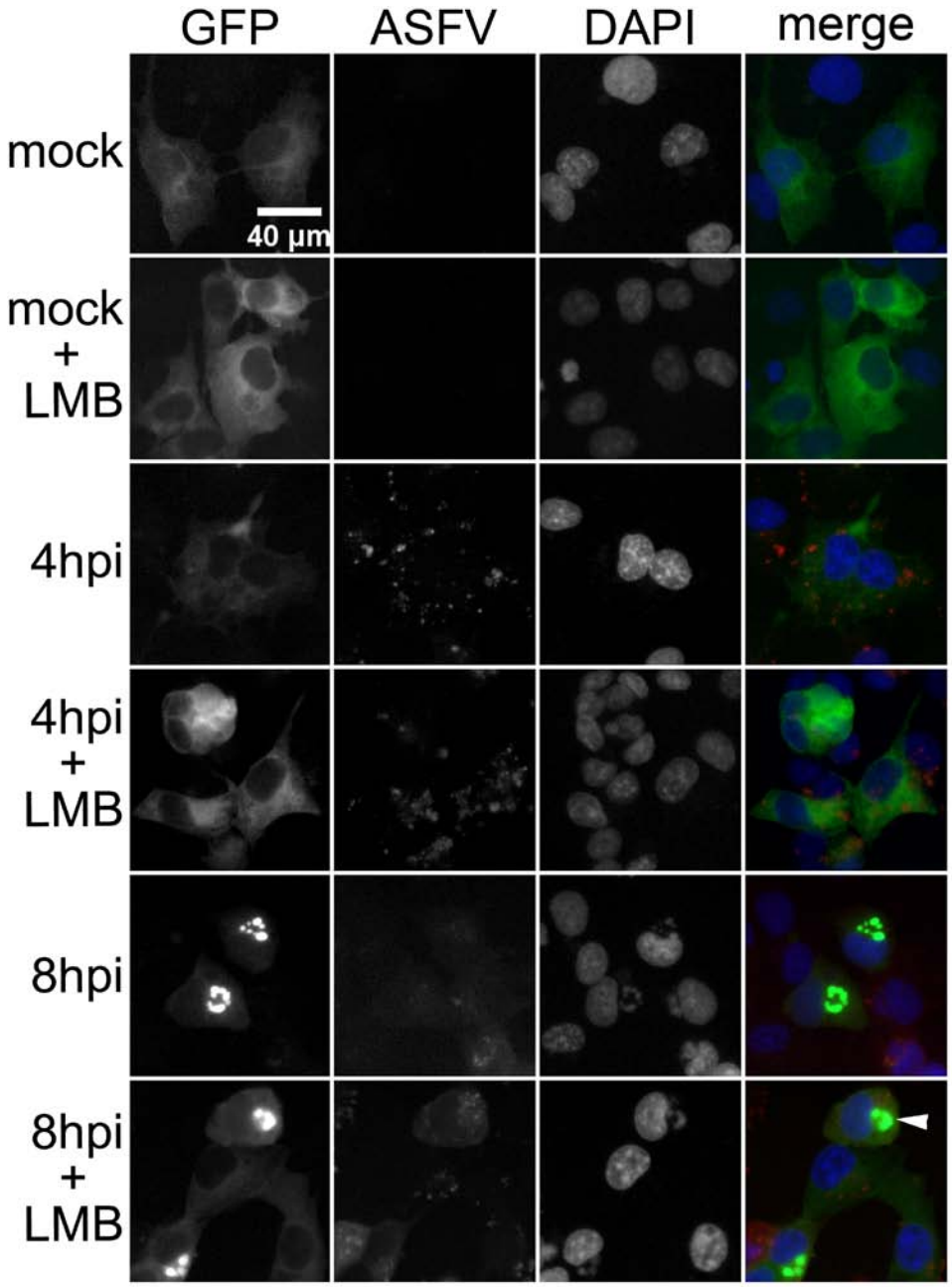


**Figure 23.** Cellular distribution of GFP-pP1192R upon viral infection.
COS-7 cells were transfected with GFP-P1192R, infected with Ba71V and analyzed at 4hpi and 8hpi. Where indicated, the cells were treated with 20 ng/ml leptomycin B (LMB) for 3 hours. In the merged panels, the GFP signal is shown in green, the anti-ASFV signal in red and DNA DAPI staining in blue. Representative microscopy images of at least two independent experiments are shown. A white arrowhead indicates an example of a viral factory.

### 3.3.2 Immunofluorescence analysis of viral pP1192R localization and expression

The results observed in COS-7 cells transfected with GFP-pP1192R and infected with Ba71V prompted us to analyze the cellular localization of viral pP1192R by immunofluorescence, using an in-house antiserum raised against a full-length recombinant pP1192R, described in detail in Chapter 4. This antiserum was shown to recognize GFP-pP1192R in COS-7 transfected cells (Figure 24).

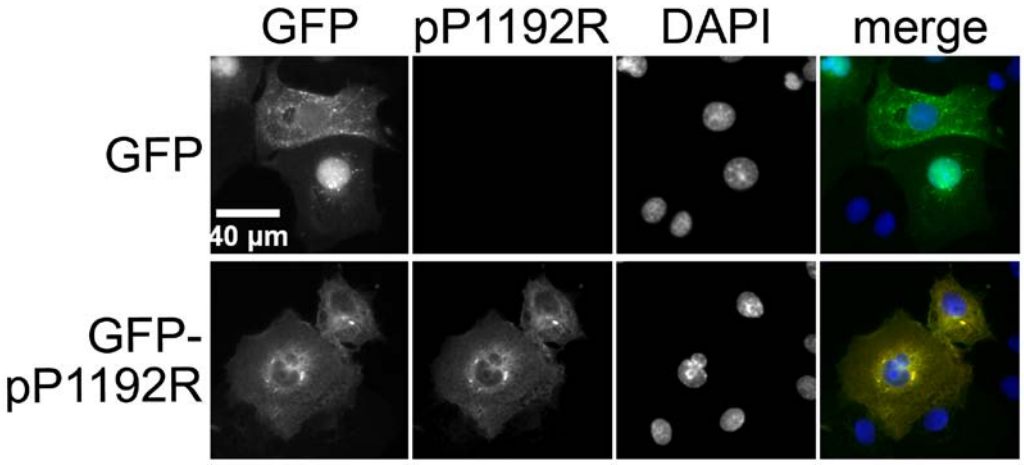


**Figure 24.** Recognition of recombinant pP1192R by the anti-pP1192R serum.
COS-7 cells were transfected with GFP or GFP-P1192R and analyzed by immunofluorescence. In the merged panels, the GFP-P1192R signal is shown in green, the anti-pP1192R signal in red and DNA DAPI staining in blue.

In an infection time-course using Ba71V-infected Vero cells pP1192R was detected as early as 6hpi, being present in several cytoplasmic foci (Figure 25). At later times the protein co-localized with and accumulated in viral factories, and was never observed in the nucleus of the infected cells. Similar results were observed in pig macrophages infected with L60 (Figure 26), although, in this case, pP1192R was only detected after 8hpi, already co-localizing with viral factories, where it accumulated over the course of the infection.

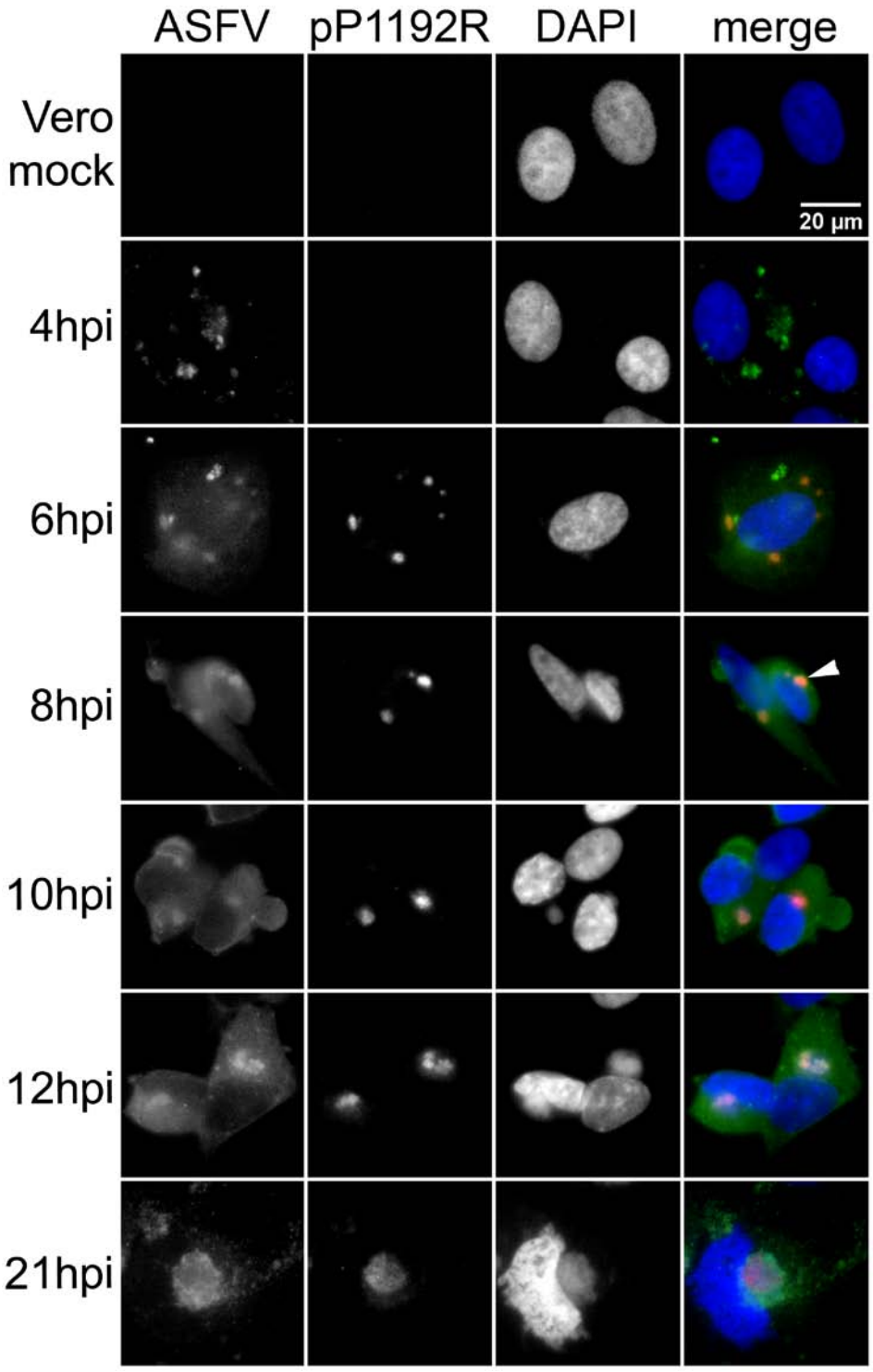


**Figure 25.** Cellular distribution of viral pP1192R during Ba71V infection in Vero cells.
Vero cells were infected with Ba71V and analyzed by immunofluorescence at the indicated time-points. In the merged panels the anti-ASFV signal is shown in green, the anti-pP1192R signal in red and DNA DAPI staining in blue. Representative microscopy images of at least two independent experiments are shown. The white arrowhead indicates an example of a viral factory.

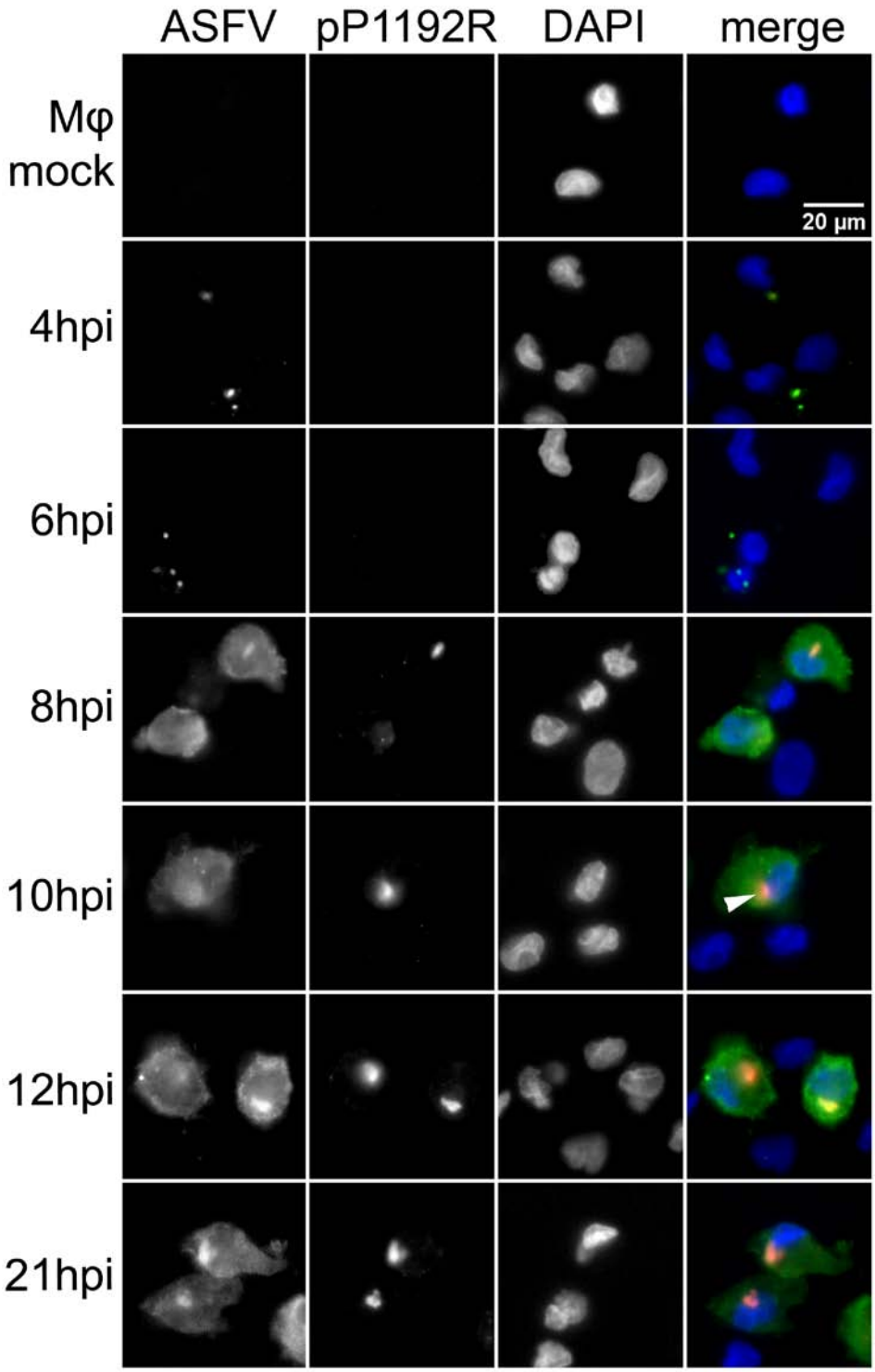


**Figure 26.** Cellular distribution of viral pP1192R during L60 infection in pig macrophages.
Pig macrophages (Mφ) were infected with L60 and analyzed by immunofluorescence at the indicated time-points. In the merged panels the anti-ASFV signal is shown in green, the anti-pP1192R signal in red and DNA DAPI staining in blue. Representative microscopy images of at least two independent experiments are shown. The white arrowhead indicates an example of a viral factory.

These results are in accordance with García-Beato *et al.* (1992) who state that P1192R RNA is mainly detected in the late phase of infection. However, these authors also detected vestigial amounts of P1192R RNA after an 18-hour infection in the presence of cytosine arabinoside (AraC), which inhibits viral DNA replication, suggesting that P1192R is transcribed, even if at a very low level, in the early phase of infection. Consequently, we performed immunofluorescence studies in the presence of AraC using Ba71V-infected Vero cells. At 4hpi no difference is observed between AraC untreated and treated infected-cells (Figure 27), with no detectable signs of pP1192R, while at 16hpi pP1192R was clearly detected in untreated infected-cells but remained undetectable in AraC-treated infected cells.

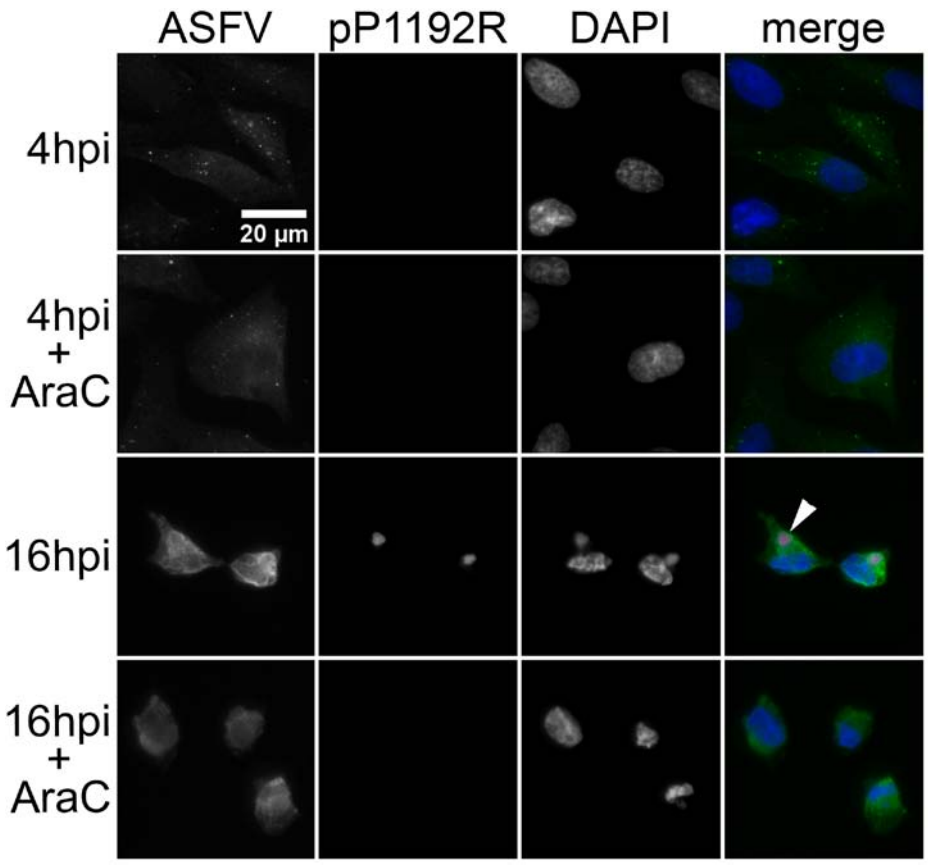


**Figure 27.** Expression of P1192R in the presence of AraC.
Vero cells were infected with Ba71V in the presence of 50 µg/ml of AraC and analyzed by immunofluorescence at the indicated time-points. In the merged panels, the anti-ASFV signal is shown in green, the anti-P1192R signal in red and DNA DAPI staining in blue. Representative microscopy images of three independent experiments are shown. The white arrowhead indicates an example of a viral factory.

The effect of AraC upon Ba71V infection was monitored by expression of a typical early ASFV protein, vp32, and of a late protein, vp72. While in both untreated and treated

Ba71V-infected cells vp32 was always detected (Figure 28), expression of vp72 was never detected upon AraC treatment of infected cells (Figure 29), confirming the effectiveness of the drug.

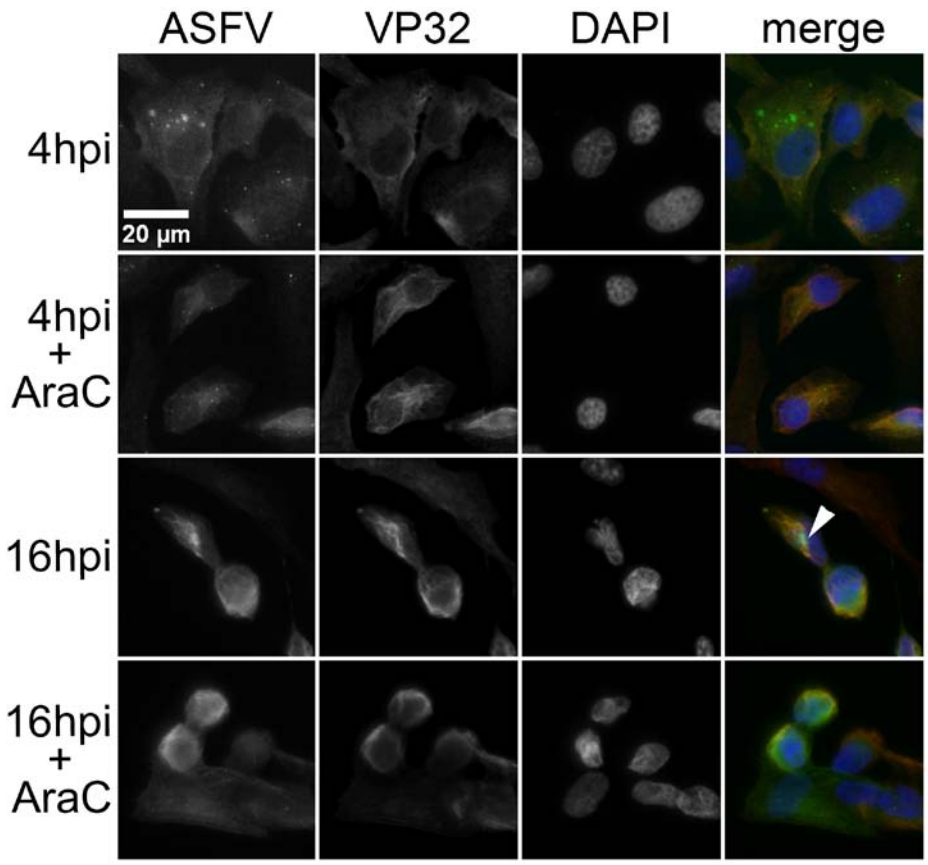


**Figure 28.** Expression of vp32 in the presence of AraC.
Vero cells were infected with Ba71V in the presence of 50 µg/ml of AraC and analyzed by immunofluorescence at the indicated time-points. In the merged panels, the anti-ASFV signal is shown in green, the anti-vp32 signal in red and DNA DAPI staining in blue. Representative microscopy images of three independent experiments are shown. The white arrowhead indicates an example of a viral factory.

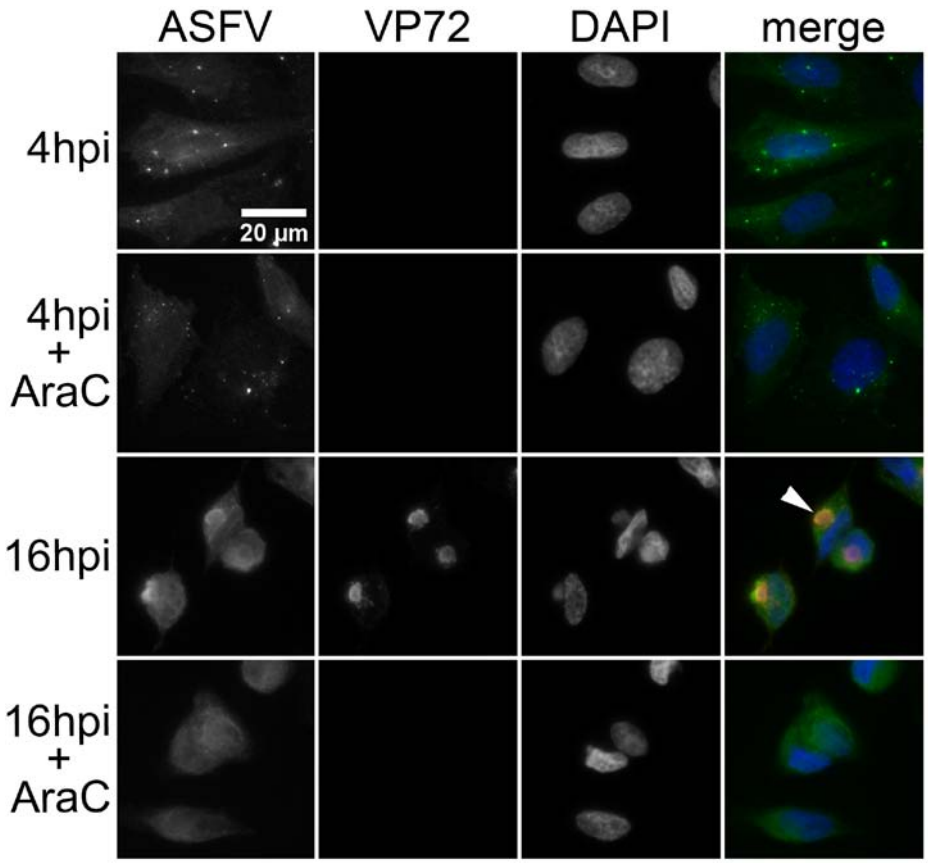


**Figure 29.** Expression of vp72 in the presence of AraC.
Vero cells were infected with Ba71V in the presence of 50 µg/ml of AraC and analyzed by immunofluorescence at the indicated time-points. In the merged panels, the anti-ASFV signal is shown in green, the anti-vp72 signal in red and DNA DAPI staining in blue. Representative microscopy images of three independent experiments are shown. The white arrowhead indicates an example of a viral factory.

Topoisomerase II poisons are drugs that trap the enzyme bound to the DNA, after generation of the double-strand cut, leading to increased levels of cleaved complexes (Nitiss, 2009). Fluoroquinolones are topoII-targeting drugs frequently used to treat bacterial infections by acting as topoII poisons (Aldred, Kerns, & Osheroff, 2014). Recently, it was described that fluoroquinolones interfere with ASFV infection (Mottola *et al.*, 2013) and one of their targets may very well be pP1192R. Therefore, we sought to use these drugs, namely enrofloxacin (a third generation fluoroquinolone), in order to trap pP1192R onto viral DNA and confirm if the anti-viral effect could be, even if partially, due to inhibition of the viral topoisomerase. Additionally, this would be a demonstration of the functionality of pP1192R as a type II DNA topoisomerase. To accomplish this, Vero cells were infected with Ba71V and incubated with or without enrofloxacin. This was followed by an adaptation of a protocol developed for the analysis of the differential retention of topoII in mammalian cells treated with topoisomerase II poisons (Agostinho *et al.*, 2004), which relies on a mild-detergent extraction prior to cell

fixation. However, in several attempts made the extraction procedure always led to the loss of the infected cells, making it impossible to gain any knowledge from this assay.

### 3.3.3 Analysis of viral pP1192R by Western blot

Contrary to what was accomplished by immunofluorescence, detection of expression of pP1192R during ASFV infection by Western blot, using the same anti-pP1192R anti-serum, was troublesome. Still, we were able to perform a time-course of infection in which a faint band of the expected molecular size (around 135 kDa) was detected, absent from extracts of mock-infected cells and present in protein extracts of Ba71V-infected Vero cells, starting from 12 hpi (Figure 30).

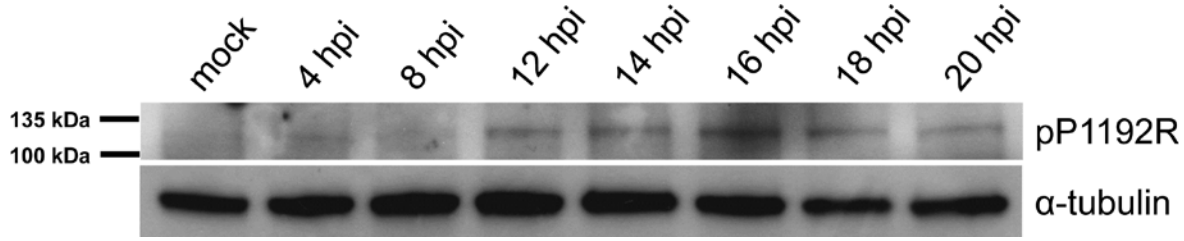


**Figure 30.** Detection of viral pP1192R by Western blot.
Protein extracts of Ba71V-infected Vero cells were analyzed by immunoblotting at the indicated time-points. Cellular α-tubulin was used as a loading control.

## 3.4 Discussion

We have found that ASFV ORF P1192R is expressed in infected cells, originating a protein with about the expected molecular weight that is localized in the cytoplasm and accumulates in viral factories over the course of infection. Although P1192R was identified as having homology to type II topoisomerases (Baylis, Dixon, Vydelingum, & Smith, 1992; García-Beato *et al.*, 1992) and it has been suggested that type II topoisomerase-inhibiting drugs have an effect on ASFV (Salas, Kuznar, & Viñuela, 1983; Mottola *et al.*, 2013), direct proof of pP1192R functionality is yet to be obtained and its precise role during viral infection remains undetermined. The retention of pP1192R onto viral DNA using fluoroquinolones could have hinted for activity, but unfortunately the assay did not work. One explanation for this may be that the protocol was initially developed for nuclear topoisomerases, which may require more harsh extraction conditions than a cytoplasmic one. We tried to lighten the procedure, without success. The cellular changes induced by ASFV infection must also

contribute to cell detachment and loss during the extraction, since this problem wasn't observed in mock-infected cells.

Previous results (García-Beato *et al.*, 1992) suggested that ORF P1192R is transcribed prior to viral DNA replication, which defines the onset of the late phase of infection. AraC, by inhibiting DNA replication, traps ASFV infection in its early phase, and no expression of late proteins is observed while many early-expressed proteins tend to accumulate over time (Santarén & Viñuela, 1986). Therefore, if the detected P1192R mRNAs were translated at early infection times, treatment of infected cells with AraC would allow us to observe viral pP1192R. However, we could not detect this protein in the presence of AraC and, therefore, it is possible that it does not participate in early replication steps, although a role in transcription of late-expressed genes cannot be ruled out.

## 3.5 References


Adachi, N., Miyaike, M., Kato, S., Kanamaru, R., Koyama, H., & Kikuchi, A. (1997). Cellular distribution of mammalian DNA topoisomerase II is determined by its catalytically dispensable C-terminal domain. *Nucleic Acids Research*, *25*(15), 3135–3142.

Agostinho, M., Rino, J., Braga, J., Ferreira, F., Steffensen, S., & Ferreira, J. (2004). Human topoisomerase IIalpha: targeting to subchromosomal sites of activity during interphase and mitosis. *Molecular Biology of the Cell*, *15*(5), 2388–2400.

Aldred, K. J., Kerns, R. J., & Osheroff, N. (2014). Mechanism of Quinolone Action and Resistance. *Biochemistry*, *53*(10), 1565–1574.

Alsner, J., Svejstrup, J. Q., Kjeldsen, E., Sørensen, B. S., & Westergaard, O. (1992). Identification of an N-terminal domain of eukaryotic DNA topoisomerase I dispensable for catalytic activity but essential for in vivo function. *The Journal of Biological Chemistry*, *267*(18), 12408–12411.

Austin, C. A., & Marsh, K. L. (1998). Eukaryotic DNA topoisomerase II beta. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *20*(3), 215–26.

Ballester, M., Galindo-Cardiel, I., Gallardo, C., Argilaguet, J. M., Segalés, J., Rodríguez, J. M., & Rodríguez, F. (2010). Intranuclear detection of African swine fever virus DNA in several cell types from formalin-fixed and paraffin-embedded tissues using a new in situ hybridisation protocol. *Journal of Virological Methods*, *168*(1-2), 38–43.

Baylis, S. A., Dixon, L. K., Vydelingum, S., & Smith, G. L. (1992). African swine fever virus encodes a gene with extensive homology to type II DNA topoisomerases. *Journal of Molecular Biology*, *228*(3), 1003–1010.

Benarroch, D., Claverie, J., Raoult, D., & Shuman, S. (2006). Characterization of mimivirus DNA topoisomerase IB suggests horizontal gene transfer between eukaryal viruses and bacteria. *Journal of Virology*, *80*(1), 314–321.

Brookes, S. M., Dixon, L. K., & Parkhouse, R. M. (1996). Assembly of African Swine fever virus: quantitative ultrastructural analysis in vitro and in vivo. *Virology*, *224*(1), 84–92.

Caeiro, F., Meireles, M., Ribeiro, G., & Costa, J. V. (1990). In vitro DNA replication by cytoplasmic extracts from cells infected with african swine fever virus. *Virology*, *179*(1), 87–94.

Caron, P. R., Watt, P., & Wang, J. C. (1994). The C-terminal domain of Saccharomyces cerevisiae DNA topoisomerase II. *Molecular and Cellular Biology*, *14*(5), 3197–207.

Cheeseman, I. M., & Desai, A. (2005). A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Science's STKE : Signal Transduction Knowledge Environment*, *2005*(266), pl1.

De León, P., Bustos, M. J., & Carrascosa, A. L. (2013). Laboratory methods to study African swine fever virus. *Virus Research*, *173*(1), 168–79.

Dickey, J. S., Choi, T.-J., Van Etten, J. L., & Osheroff, N. (2005). Chlorella virus Marburg topoisomerase II: high DNA cleavage activity as a characteristic of Chlorella virus type II enzymes. *Biochemistry*, *44*(10), 3899–908.

Engel, R., Valkov, N. I., Gump, J. L., Hazlehurst, L., Dalton, W. S., & Sullivan, D. M. (2004). The cytoplasmic trafficking of DNA topoisomerase IIα correlates with etoposide resistance in human myeloma cells. *Experimental Cell Research*, *295*(2), 421–431.

Espeli, O., Levine, C., Hassing, H., & Marians, K. J. (2003). Temporal regulation of topoisomerase IV activity in E. coli. *Molecular Cell*, *11*, 189–201.

Eulálio, A., Nunes-Correia, I., Salas, J., Salas, M. L., Simões, S., & Pedroso de Lima, M. C. (2007). African swine fever virus p37 structural protein is localized in nuclear foci containing the viral DNA at early post-infection times. *Virus Research*, *130*(1-2), 18–27.

García-Beato, R., Freije, J. M., López-Otín, C., Blasco, R., Viñuela, E., & Salas, M. L. (1992). A gene homologous to topoisomerase II in African swine fever virus. *Virology*, *188*(2), 938–47.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell*, *23*(1), 175–182.

Huang, W. M., Wei, L. S., & Casjens, S. (1985). Relationship between bacteriophage T4 and T6 DNA topoisomerases. T6 39-protein subunit is equivalent to the combined T4 39- and 60-protein subunits. *Journal of Biological Chemistry* , *260* (15 ), 8973–8977.

Jensen, S., Andersen, a H., Kjeldsen, E., Biersack, H., Olsen, E. H., Andersen, T. B., Westergaard, O., & Jakobsen, B. K. (1996). Analysis of functional domain organization in DNA topoisomerase II from humans and Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *16*(7), 3866–77.

Lavrukhin, O. V, Fortune, J. M., Wood, T. G., Burbank, D. E., Van Etten, J. L., Osheroff, N., & Lloyd, R. S. (2000). Topoisomerase II from Chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase. *The Journal of Biological Chemistry*, *275*(10), 6915–21.

Leitão, A., Cartaxeiro, C., Coelho, R., Cruz, B., Parkhouse, R. M., Portugal, F., Vigário, J. D., & Martins, C. L. (2001). The non-haemadsorbing African swine fever virus isolate ASFV/NH/P68 provides a model for defining the protective anti-virus immune response. *The Journal of General Virology*, *82*(Pt 3), 513–23.

Leitão, A., Malur, A., Cartaxeiro, C., Vasco, G., Cruz, B., Cornelis, P., & Martins, C. L. V. (2000). Bacterial lipoprotein based expression vectors as tools for the characterisation of African swine fever virus (ASFV) antigens. *Archives of Virology*, *145*(8), 1639–1657.

Liu, L. F., Liu, C.-C., & Alberts, B. M. (1979). T4 DNA topoisomerase: a new ATP-dependent enzyme essential for initiation of T4 bacteriophage DNA replication. *Nature*, *281*(5731), 456–461.

Mirski, S. E. ., Bielawski, J. C., & Cole, S. P. . (2003). Identification of functional nuclear export sequences in human topoisomerase IIα and β. *Biochemical and Biophysical Research Communications*, *306*(4), 905–911.

Mirski, S. E., Gerlach, J. H., Cummings, H. J., Zirngibl, R., Greer, P. a, & Cole, S. P. (1997). Bipartite nuclear localization signals in the C terminus of human topoisomerase II alpha. *Experimental Cell Research*, *237*(2), 452–5.

Mirski, S. E. L., Gerlach, J. H., & Cole, S. P. C. (1999). Sequence Determinants of Nuclear Localization in the α and β Isoforms of Human Topoisomerase II. *Experimental Cell Research*, *339*, 329–339.

Mo, Y. Y., Wang, C., & Beck, W. T. (2000). A novel nuclear localization signal in human DNA topoisomerase I. *The Journal of Biological Chemistry*, *275*(52), 41107–41113.

Mottola, C., Freitas, F. B., Simões, M., Martins, C., Leitão, A., & Ferreira, F. (2013). In vitro antiviral activity of fluoroquinolones against African swine fever virus. *Veterinary Microbiology*, *165*(1-2), 86–94.

Nitiss, J. L. (2009). Targeting DNA topoisomerase II in cancer chemotherapy. *Nature Reviews. Cancer*, *9*(5), 338–350.

Portugal, R., Leitão, A., & Martins, C. (2009). Characterization of African swine fever virus IAP homologue expression in porcine macrophages infected with different virulence isolates. *Veterinary Microbiology*, *139*(1-2), 140–6.

Salas, M. L., Kuznar, J., & Viñuela, E. (1983). Effect of rifamycin derivatives and coumermycin A1 on in vitro RNA synthesis by African Swine Fever Virus. *Archives of Virology*, *77*(1), 77–80.

Santarén, J. F., & Viñuela, E. (1986). African swine fever virus-induced polypeptides in Vero cells. *Virus Research*, *5*, 391–405.

Shaffer, R., & Traktman, P. (1987). Vaccinia virus encapsidates a novel topoisomerase with the properties of a eucaryotic type I enzyme. *Journal of Biological Chemistry* , *262* (19 ), 9309–9315.

Tadesse, S., & Graumann, P. L. (2006). Differential and dynamic localization of topoisomerases in Bacillus subtilis. *Journal of Bacteriology*, *188*, 3002–3011.

Wessel, I., Jensen, P. B., Falck, J., Mirski, S. E. L., Cole, S. P. C., & Sehested, M. (1997). Loss of Amino Acids 1490Lys-Ser-Lys1492 in the COOH-Terminal Region of Topoisomerase IIα in Human Small Cell Lung Cancer Cells Selected for Resistance to Etoposide Results in an Extranuclear Enzyme Localization. *Cancer Research* , *57*(20 ), 4451–4454.

Yáñez, R. J., Rodríguez, J. M., Nogal, M. L., Yuste, L., Enríquez, C., Rodriguez, J. F., & Viñuela, E. (1995). Analysis of the complete nucleotide sequence of African swine fever virus. *Virology*, *208*(1), 249–78.

# CHAPTER 4

# HETEROLOGOUS EXPRESSION AND *IN VITRO* FUNCTIONAL CHARACTERIZATION OF ASFV ORF P1192R

## 4.1 Introduction

Many strategies can be devised to study a specific protein, to determine its localization and activity in the cellular environment, and to identify other proteins that interact with it. However, many of the proteins existing inside of a cell (even if they are of viral origin) are in quantities so minute that it's extremely difficult to perform biochemical assays in order to accurately characterize them. Also, unless highly-specific antibodies are available for the protein under study, which is rare for scarcely characterized proteins, purifying it from a pool of proteins such as a cellular extract can be a daunting task. The development of systems for heterologous expression of recombinant proteins allowed researchers to overcome these handicaps. Nowadays, expression systems range from cell-free systems, such as rabbit reticulocyte lysates and *Escherichia coli* cell-free extracts (Murray & Baliga, 2013), to prokaryotic systems, with *E. coli* being the best developed and most widely used and gram-positives *Lactococcus lactis* and *Bacillus* spp. as interesting alternatives (Sugiki, Fujiwara, & Kojima, 2014), or eukaryotic systems, which have a broader range since they include fungi [yeasts like *Pichia pastoris*, *Saccharomyces cerevisiae* or *Kluyveromyces lactis* (Mattanovich *et al.*, 2012), or filamentous fungi, namely *Aspergillus* spp. or *Neurospora crassa* (Su, Schmitz, Zhang, Mackie, & Cann, 2012)], insect cells (with or without infection with baculovirus, usually *Autographa californica* multicapsid nucleopolyhedrovirus) (Drugmand, Schneider, & Agathos, 2012), mammalian culture cells (with CHO and HEK 293 cell lines leading in the production of proteins with biopharmaceutical value) (Zhu, 2012; Sugiki *et al.*, 2014) or whole organisms like transgenic plants (tobacco plant) (Bhoo *et al.*, 2011) or transgenic animals (mainly sheep) (Simons *et al.*, 1988; Houdebine, 2009). Since production of a recombinant functional protein is tightly linked to the cellular machinery of the expression system being used, the choice of system must be made factoring in any specific requirements for expression of the protein, like post-translational modifications, codon usage or protein folding, as well as the future use of the protein.

### 4.1.1 Recombinant protein expression in *Escherichia coli*

*E. coli* is frequently the first choice when picking a system for recombinant protein expression. The easiness of its genetic manipulation allowed for the fine-tuning of expression strains and the development of protein expression strategies, but there are other major advantages: it’s a fast growing organism in rich, inexpensive media, with a doubling time of 20 minutes under optimal conditions (Sezonov, Joseleau-Petit, & D’Ari, 2007), able to reach high cell densities in culture; it’s easily and rapidly transformed with exogenous DNA; it expresses recombinant proteins at very high levels, resulting in elevated yields of purified

protein. Still, while the burden of protein expression usually impacts on the growth rate, diminishing or nearly abolishing it, this is usually taken into account when considering heterologous expression in this system. It also has the disadvantage of lacking the capacity to perform post-transcriptional modifications often required for proper activity of eukaryotic proteins.

Many strains of *E. coli* have been generated for the purpose of heterologous protein expression. Of these, BL21(DE3) (Daegelen, Studier, Lenski, Cure, & Kim, 2009) is one of the most commonly used. It was derived from strain BL21 by insertion, in the bacterial chromosome, of the λDE3 prophage carrying an inducible gene coding for the T7 RNA polymerase (RNApol) under the control of the *lac*UV5 promoter (Studier & Moffatt, 1986). Other relevant genetic traits in BL21(DE3) include deletion of the outer membrane protease OmpT, which could contribute to protein degradation after cell lysis (Grodberg & Dunn, 1988), and deletion of the Lon protease, which is one of the proteases responsible for the degradation of abnormally folded proteins in *E. coli* (Gottesman, 1996). The T7 RNApol coded by the λDE3 prophage will actively transcribe genes placed under control of the T7 promoter at much higher rates than the endogenous *E. coli* RNApol would, thereby boosting recombinant protein expression. This implies that the ORF of interest must be cloned downstream of a T7 promoter. The *lac*UV5 promoter controlling the T7 RNApol can be induced through addition to the culture medium of lactose or of a non-hydrolyzable analog like isopropyl-β-D-1-thiogalactopyranoside (IPTG), permitting temporal control of protein expression (Figure 31). It has, however, the disadvantage of being a “leaky” promoter, i.e., expression from this promoter is detected even in the absence of an inducer. Furthermore, the *lac*UV5 promoter is a modified version of *lac* promoter, selected for its reduced sensitivity to catabolite regulation (Silverstone, Arditti, & Magasanik, 1970), which makes it more “leaky” than its wild-type version. This means that even with glucose in the culture medium, which for the wild-type *lac* promoter extremely reduces transcription of genes under its control (Donovan, Robinson, & Glick, 1996), expression of *lac*UV5-controled genes will occur. This can be deleterious for the cell, if the recombinant protein being expressed is somehow toxic, but may be counteracted by co-expressing the T7 lysozyme, which inhibits initiation of transcription from the T7 promoter through binding to T7 RNApol (Moffatt & Studier, 1987). T7 lysozyme is usually supplied in a plasmid, pLysS or pLysE, with which BL21(DE3) cells are transformed (Studier, 1991). These two plasmids differ in the amount of T7 lysozyme produced by the cell, with pLysE allowing for production of higher amounts of the enzyme. In BL21(DE3) cells containing either plasmid, and in the absence of induction of protein expression, the T7 lysozyme levels suffice to inhibit the T7 RNApol being expressed due to

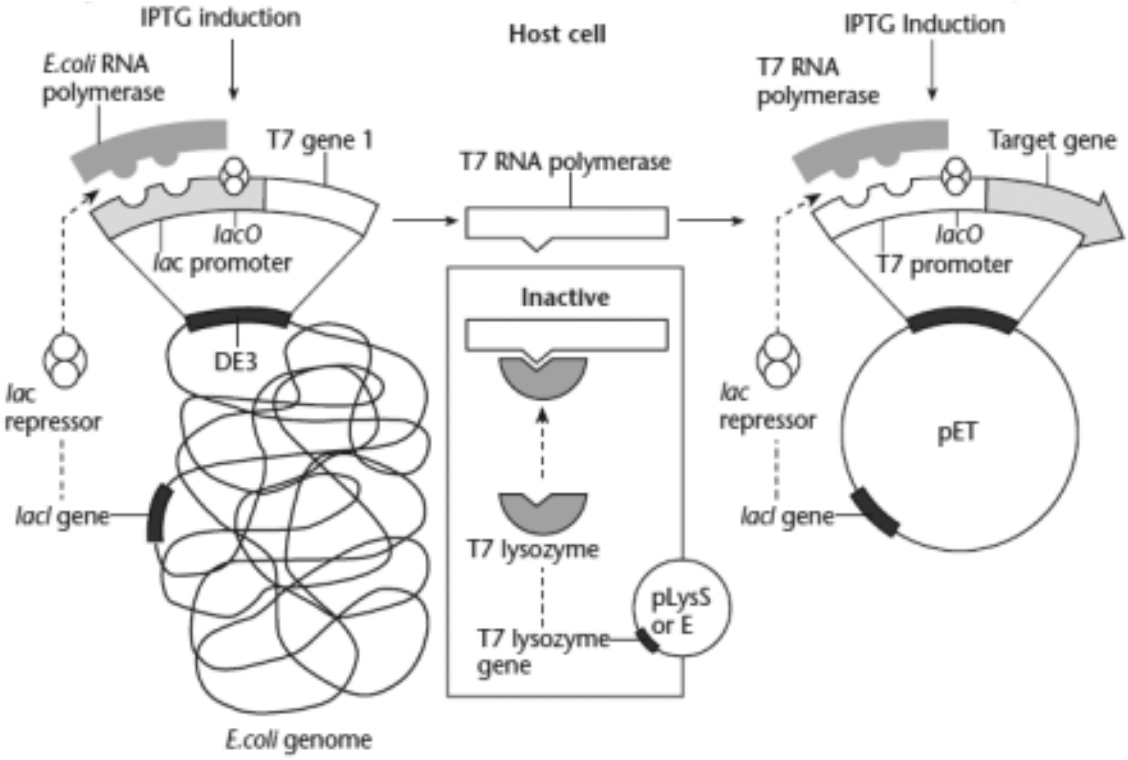


**Figure 31.** Regulation of protein expression in the pET system.
Figure was adapted from the pET system manual (Novagen).

"leakiness" of its promoter. Upon addition of lactose or IPTG, expression of T7 RNApol will increase and surpass the levels of T7 lysozyme, which will not be sufficient to sustain inhibition. In this way, expression of any foreign gene placed under the control of the T7 promoter will be tightly regulated and dependent of induction with an inducer of the *lac* promoter.

This was the strategy adopted during our work, but a myriad of other *E. coli*-based expression systems is available.

### 4.1.2 Recombinant protein expression in yeast

Unicellular eukaryotic organisms like yeasts are an excellent alternative to *E. coli* for heterologous protein expression, namely due to their ability to perform accurate folding of complex proteins and post-translational modifications, often necessary during expression of eukaryotic proteins. In fact, the yeast *Saccharomyces cerevisiae* is almost the eukaryotic equivalent of *E. coli*. It's a unicellular organism, very-well studied and for which a plethora of tools and knowledge have been developed. Its genome has been fully sequenced (Goffeau *et al.*, 1996) and its genetic manipulation is straightforward. It is easily grown in culture, using simple or complex culture medium, up to high cell densities. Being an eukaryotic organism

with a sub-cellular organization similar to higher eukaryotes makes it suitable for heterologous expression of cytosolic (and usually soluble) or secreted proteins, especially since its use is less expensive than that of insect or mammalian cells. For the past few decades *S. cerevisiae* has been used for research, industrial or medical purposes, and as an organism classified as "generally recognized as safe" it is viewed as a viable option for expression of proteins with pharmaceutical applications. In recent years other yeasts, like *Pichia pastoris* or *Kluyveromyces lactis*, have also gained interest due to some of their intrinsic properties, like promoter usage or high-level secretion. Indeed, yeast systems are often adopted for production of proteins that are naturally secreted. This happens not only because of the efficiency of their secretion apparatus but also because yeasts, in general, secrete only a minute fraction of endogenous proteins. This offers a means to avoid eventual toxicity induced by expression of a foreign protein and also simplifies the purification procedure. One possible downside resides in the patterns of glycosylation of the secreted proteins, since these differ from what occurs in higher eukaryotes. While eukaryotic proteins have sialylated O-linked chains, in *S. cerevisiae* O-linked oligosaccharides contain only mannose moieties and N-linked sites are frequently hyperglycosylated (Romanos, Scorer, & Clare, 1992). This difference can be of relevance when expressing proteins for therapeutic applications, since yeast-like glycosylation may, for example, trigger an undesired immune response or reduce the efficacy of therapy (Li & d'Anjou, 2009).

Based on their mode of replication, *S. cerevisiae* vectors are classified in five distinct classes: integrative (YIp), episomal (YEp), replicating (YRp), centromeric (YCp) and linear (YLp) (Lundblad, 2001). While integrative vectors lack sequences required for autonomous replication and have to be integrated into the genome for propagation, the linear vectors contain guanine-rich repeats at their termini that act like telomeres and allow the molecule to replicate and be maintained in a linear form. On the other hand, episomal, replicating and centromeric vectors are circular and replicate independently of the yeast chromosomes, but they differ in their copy number and in the origin of the sequences that confer them replicative autonomy. YEp vectors are derived from the naturally occurring 2µm plasmid and are present in very high copy numbers (20-50 copies per cell). YCp vectors contain segments from yeast centromeres, behaving like yeast chromosomes and being present in very low copy numbers (1-2 copies per cell). Finally, the YRp vectors possess autonomous replication sequences (ARS), which are no more than chromosomal origins of replication. These ARS sequences permit the YRp vectors to exist in elevated copy numbers (up to 100 copies per cell) but they are very unstable, with high rates of plasmid loss during mitosis and meiosis due to inefficient segregation to daughter cells. For protein expression purposes, the most advantageous are the

YEp vectors, which allow for elevated copy numbers per cell of the gene of interest. Contrary to what happens in *E. coli*, in which selection of transformants is usually performed through the use of antibiotics, selection in *S. cerevisiae* is typically achieved using auxotrophic markers (Romanos *et al.*, 1992). Therefore, continued selection requires the use of defined media lacking the corresponding nutrient. The most common selection markers for *S. cerevisiae* transformants are *URA3*, *HIS3*, *LEU2* and *TRP1*, that can be used in corresponding mutants strains auxotrophic for uracil, histidine, leucine or tryptophan. *URA3* and *TRP1* have the additional advantage of being compatible with the use of protein hydrolysates (for example, casamino acids) in the culture medium, therefore allowing the use of semi-defined medium in order to promote higher growth rates. Still, antibiotics can also be used for selection, and good examples are selection with G418 (Davies & Jimenez, 1980) and zeocin (Gatignol, Baron, & Tiraby, 1987).

The choice of promoters for foreign protein expression in *S. cerevisiae* is also a hard one, as there are many available hypotheses (Romanos *et al.*, 1992). In resemblance to *E. coli*, preference goes to inducible promoters, thereby having the possibility to avoid detrimental effects of constitutive protein expression (for example, due to toxicity of the recombinant protein). The most powerful, tightly-regulated promoters of *S. cerevisiae* are those of genes involved in galactose metabolism, like *GAL1*, *GAL7* and *GAL10* (St John & Davis, 1981). The mechanism of induction by galactose is broadly similar to that of lactose induction in *E. coli*, and is also repressed by the presence of glucose in the culture medium, so that induction only starts following glucose depletion or removal. Growth in the presence of a non-repressing sugar, like raffinose, is also an option, with induction being started just by addition of galactose to the culture medium.

Next to *S. cerevisiae*, the *P. pastoris* system is the most developed for heterologous protein expression. It presents several advantages over baker's yeast: (i) *P. pastoris* avoids using carbon sources for fermentation and prefers respiratory growth (Higgins & Cregg, 1998), thereby avoiding the generation of toxic byproducts of fermentation, such as ethanol, allowing cultures to reach even higher cell densities; (ii) glycosylation in *P. pastoris* is similar to that of higher eukaryotes, with oligosaccharides having shorter chain lengths than in *S. cerevisiae* (Bretthauer & Castellino, 1999), and a strain has been developed in order to produce "humanized" glycoproteins (Hamilton *et al.*, 2006); (iii) *P. pastoris* is able to utilize methanol as a sole carbon and energy source, since it possesses a set of enzymes dedicated to this metabolic pathway (Yurimoto, Oku, & Sakai, 2011). Two of these enzymes, the alcohol oxidase and the dihydroxyacetone synthase, can make up more than 30% of the total protein content of cells cultured in the presence of methanol (Couderc & Baratti, 1980), even though

they are absent during growth based on glucose, and their regulation is mainly at the transcriptional level (Tschopp, Brust, Cregg, Stillman, & Gingeras, 1987). It is therefore no surprise that the promoter of *AOX1*, the gene coding for the predominant alcohol oxidase (Cregg, Madden, Barringer, Thill, & Stillman, 1989), is usually the promoter of choice for recombinant protein expression in *P. pastoris*, as it will allow high-level protein production in the presence of methanol. Because transformation of this yeast commonly involves integration of the vector, as it has no natural plasmids (Romanos *et al.*, 1992), and since integration is generally directed to the *AOX1* promoter region, by homologous recombination, integration at this site may disrupt the promoter and lead to loss of the majority of cellular alcohol oxidase activity. The ability of a strain of *P. pastoris* to utilize methanol as the sole carbon source is referred to as the Mut phenotype and is dependent on the activity of the alcohol oxidase. Consequently, cells in which the *AOX1* promoter is disrupted are denoted as having a $Mut^S$ (for methanol utilization slow) phenotype because they grow poorly on methanol-containing medium. They rely only on the Aox2 enzyme, which is very similar to Aox1 (they share more than 90% homology) but expressed at a much lower level (Cregg *et al.*, 1989). Integration of the expression plasmid in the *P. pastoris* genome can also be useful for increasing the expression of a specific gene, since the number of integrated copies can be increased (Scorer, Clare, McCombie, Romanos, & Sreekrishna, 1994). However, this strategy doesn't always lead to an increased amount of protein being produced (Thill *et al.*, 1990).

### 4.1.3 *In vitro* assays for evaluation of topoisomerase activity

DNA topoisomerases are enzymes that change the landscape of DNA molecules by promoting catenation/decatenation, knotting/unknotting and relaxation/supercoiling. The mechanisms by which topoisomerases act involve the generation of single (for type I topoisomerases) or double (for type II topoisomerases) stranded breaks in the DNA. Agents that block theses mechanisms and inhibit topoisomerase activity or that trap topoisomerase-cleaved DNA complexes and promote the accumulation of breaks in the DNA are often used for anti-cancer and anti-bacterial treatments (Fortune & Osheroff, 2000; Anderson & Osheroff, 2001; Nitiss, 2009; Pommier, Leo, Zhang, & Marchand, 2010).

Given the importance of topoisomerases for biomedical research, simple *in vitro* assays have been developed to assess the activity of topoisomerases, either in complex protein extracts or in purified form (Nitiss, Soans, Rogojina, Seth, & Mishina, 2012), and to study their inhibition. Since type I and type II topoisomerases act through different mechanisms and have different functional requirements, their activity can be evaluated differently.

Type I topoisomerase activity can be observed through a relaxation assay, in which a supercoiled plasmid DNA becomes relaxed by action of the enzyme. Since a relaxed form of the plasmid has different electrophoretic mobility than a supercoiled form, these can be distinguished by simply migrating the reaction products in an agarose gel (Figure 32A). When assaying for activity using protein extracts, omission of ATP from the reaction buffer will prevent any topoII present in the extract from functioning and any observed activity will originate from topoI. Thus, ATP-dependency allows for a clear distinction between the two types of enzymes. The relaxation assay can also be used to assay for topoII activity, but in this case ATP and $MgCl_2$ must be included in the reaction buffer, and only purified enzyme can be used to avoid topoI contamination. For specific evaluation of type II topoisomerase activity, a decatenation assay must be used. In this assay, kinetoplast DNA (kDNA) from *Crithidia fasciculata* is used, since it forms a large network of catenated minicircles of DNA. In this form the catenated circles are unable to enter and migrate in an agarose gel, whereas upon decatenation the free circles migrate and can be observed in two distinct forms, either open circular DNA or covalently closed circular DNA (Figure 32B). Despite requiring the presence of ATP in the reaction buffer, this assay can be used to evaluate topoII activity from protein extracts since any existing topoI enzyme will not be able to decatenate the kDNA.

In both assays, the result is observed by visualization of the reaction products in an agarose gel. Ethidium bromide, like other DNA dyes, when present during the run will intercalate

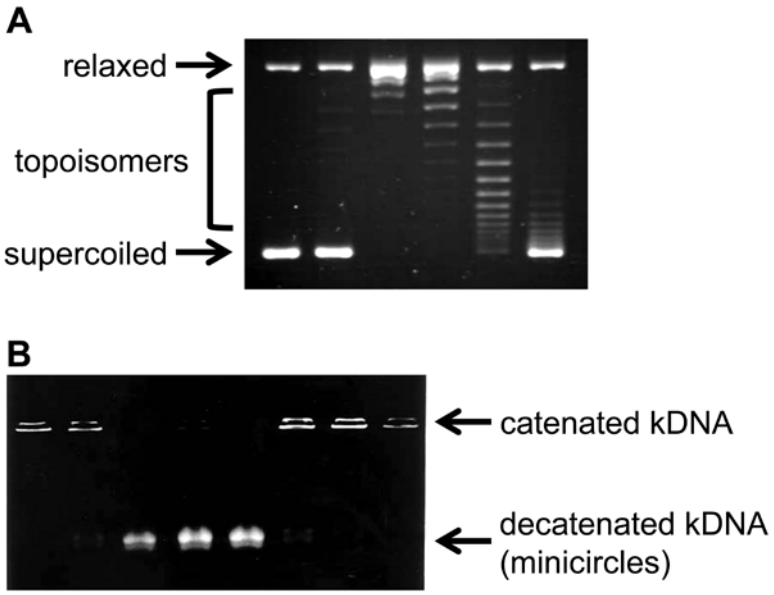


**Figure 32.** Topoisomerase activity assays.
Panel A illustrates a plasmid relaxation assay, which may be used to assess activity for type I or type II topoisomerases. It was adapted from Lavrukhin *et al.* (2000). Panel B illustrates a decatenation assay, specific for type II topoisomerases, and was adapted from Haldane and Sullivan (2001).

DNA during its migration. This intercalation induces unwinding of circular DNA molecules, eventually promoting positive supercoiling, and may thus lead to erroneous interpretations of the results (Bailly, 2001). Therefore, for visualization of reaction products from topoisomerase assays, these must be run in the absence of a DNA dye and the DNA molecules must be stained post-run.

Despite having been described more than 20 years ago (Baylis *et al.*, 1992; García-Beato *et al.*, 1992), ASFV ORF P1192R was never shown to actually code for a functional type II DNA topoisomerase. Since the work described in previous chapters reinforces the hypothesis that P1192R is indeed a functional topoisomerase II, the objective of the work described in this chapter was to demonstrate its activity, and this was done by heterologously expressing P1192R and characterizing the protein's activity *in vitro*.

## 4.2 Materials and methods

Unless stated, procedures for DNA amplification by PCR, agarose gel electrophoresis, DNA purification from agarose slices, plasmid DNA purification and DNA quantification were performed as described on Chapter 2.

### 4.2.1 Primers used

The primers used in this chapter are listed below in Table 7. Sequences of restriction sites included in the primers are underlined. Nucleotides included upstream of those sequences were added to increase efficiency of the hydrolysis reactions. Primers used were synthesized either by Life Technologies or by STAB VIDA.

**Table 7.** Primers used in this chapter.

| Primer name | Sequence (5' → 3') | Purpose |
|---|---|---|
| eGFP_KpnI_FW | CGGGGTACCATGGTGAGCAAGGGCGAGGAGC | cloning |
| eGFP_KpnI_RV | CGGGGTACCCTTGTACAGCTCGTCCATGCCG | cloning |
| GST-NotI FW | AAGGAAAAAAGCGGCCGCCATGTCCCCTATACTAGGTTATTGG | cloning |
| GST-NotI RV | AAGGAAAAAAGCGGCCGCTCAATCCGATTTTGGAGGATGGTCGC | cloning |
| P1192R ATG NdeI Up | CCCATATGATGGAAGCGTTTGAAATCAGC | cloning |
| P1192R noStop NotI Low | CCGCGGCCGCATGAAATTTCCACTGAGTA | cloning |
| P1192R pPIC-pYES RV | AAGGAAAAAAGCGGCCGCCAGTCATGTCTAAGGCTACAAACTC | cloning |
| P1192R_KpnI_ATG_FW | GGGGTACCATGGAAGCGTTTGAAATCAGCGATTTC | cloning |
| P1192R_KpnI_ATG_FW_pICZαC | GGGGTACCGATGGAAGCGTTTGAAATCAGCGATTTC | cloning |
| P1192R_noStop_NotI_RV_JC | AAGGAAAAAAGCGGCCGCCATGAAATTTCCACTGAGTATTTCTTCC | cloning |
| P1192R_NotI_Stop_RV | AAGGAAAAAAGCGGCCGCTTAATGAAATTTCCACTGAGTATTTC | cloning |
| TOP2_Cterm_FW_NotI | AAGGAAAAAAGCGGCCGCTCCCAATAAAGGGAGCAAAACGAAAGG | cloning |
| TOP2_Cterm_RV_NotI | AAGGAAAAAAGCGGCCGCTCAATCCTCTTCATTGAACGAAACATC | cloning |

Table 7 (continuation)

| Primer name | Sequence (5' → 3') | Purpose |
|---|---|---|
| TOP2_FW_BamHI | CCGGGATCCGATGTCAACTGAACCGGTAAGCGC | cloning |
| TOP2_RV_NotI_pYES | AAGGAAAAAAGCGGCCGCGGGCTTTGTTAGCAGCCGGATC | cloning |
| TOP2_RV_XhoI_pET | CCCGCTCGAGATCCTCTTCATTGAACGAAACATCTG | cloning |
| 3'-AOX1_seq | GCAAATGGCATTCTGACATCC | sequencing |
| 5'-AOX1_seq | GACTGGTTCCAATTGACAAGC | sequencing |
| CMV Fwd Primer | CGCAAATGGGCGGTAGGCGTG | sequencing |
| eGFP N Rv | CGTCGCCGTCCAGCTCGACCAG | sequencing |
| eYFP N1 Rv | AAACCTCTACAAATGTGGTATGG | sequencing |
| P1192R 2nd Frag FW | CCGAATTCCCCAATTACAATTTCAAGGGCCATTTGAAGCGGTTTGGCCA | sequencing |
| P1192R iSeq1 Up | ACAAATCAACCAGCGCCT | sequencing |
| P1192R iSeq3 Up | CGACGAATTCCTTGCAGCCT | sequencing |
| P1192R iSeq4 Up | CAACAGACGATTAAAGATAAAAACC | sequencing |
| P1192R iSeq5 Up | GAGCATGGATTTCCCCCGCTG | sequencing |
| P1192R iSeq6 Low | CTATGGCAAGCTCTTTCATGATCC | sequencing |
| P1192R_seq1 | GAAGCGTTTGAAATCAGC | sequencing |
| P1192R_seq2 | TCATTGAGCCTCCTACCA | sequencing |
| P1192R_seq3 | AAAACGCATTACTCCATCC | sequencing |
| P1192R_seq4 | TGTACGAAAAGGGCAAGA | sequencing |
| P1192R_seq5 | CCTCCCACCTCTATCCAG | sequencing |
| P1192R_seq6 | TGGTGAAACTAAAGCCAAA | sequencing |
| P1192R_seq7 | CTTAATGGTCGCCGTATG | sequencing |
| pUC/M13 FW | CGCCAGGGTTTTCCCAGTCACGAC | sequencing |
| pUC/M13 RV | TCACACAGGAAACAGCTATGAC | sequencing |
| RtPr2 P1192R Up | AGCGAGCAAAGCTGAGATACG | sequencing |
| T7_RV_JC | GCTAGTTATTGCTCAGCGG | sequencing |
| T7_FW_JC | TAATACGACTCACTATAGGG | sequencing |
| Y800F | GCAGGATCCCCAAGATTCATCAGTGTGCAGCTT | mutagenesis |
| Y800F_antisense | AAGCTGCACACTGATGAATCTTGGGGATCCTGC | mutagenesis |

Sequences recognized by restriction enzymes are shown underlined – GGTACC for *Kpn*I, GCGGCCGC for *Not*I, CATATG for *Nde*I, GGATCC for *Bam*HI, CTCGAG for *Xho*I and GAATTC for *Eco*RI. Some of the primers used for sequencing were also used for colony PCR.

### 4.2.2 Hydrolysis and ligation of DNA fragments

Restriction enzymes used for hydrolysis of plasmids and/or DNA fragments were *Age*I, *Kpn*I, *Nde*I, *Not*I, *Pme*I, *Xho*I (all from New England Biolabs) and *Bam*HI (Promega).

When cloning in plasmids hydrolyzed by a single restriction enzyme their 5' extremities were dephosphorylated to avoid their recircularization. For this, Antartic Phosphatase (AP) (New England Biolabs) was used. Following hydrolysis with the restriction enzyme in a total volume of 20 µl, to the same tube(s) 3 µl of 10X Antartic Phosphatase Reaction Buffer, 5 U

of AP and 6 µl of $H_2O$ were added, for a total of 30 µl. Reactions were incubated at 37 ºC for 1 hour, followed by 5 minutes at 70 ºC to inactive the phosphatase.

### 4.2.3 Cloning of P1192R from ASFV/L60

For expression in *E. coli*, ORF P1192R was PCR-amplified skipping the endogenous stop codon, hydrolyzed with *Nde*I and *Not*I and cloned in vector pET24a, previously hydrolyzed with the same restriction enzymes, in-frame with the existent 3' polyhistidine (6xHis) tag. This originated plasmid pET24a-P1192R_6xHis. Cloning in pET24a puts the ORF of interest under the control of the phage T7 promoter, which is not recognized by the *E. coli* RNA polymerase, permitting a very tight control over heterologous expression. Transcription is driven by the T7 RNA polymerase, which must be expressed in the host bacterial cells (either a λDE3 lysogen or infected with λCE6).

For expression in *Pichia pastoris*, ORF P1192R was cloned in plasmid pPICZαC (Life Technologies), kindly provided by Dr. Rodrigo Cunha, Pelotas, Brazil. This plasmid contains, upstream of the multiple cloning site (MCS), a sequence coding for the *S. cerevisiae* α-factor secretion signal, for secretion of the recombinant protein, and downstream of the MCS, sequences coding for the *c-myc* epitope and for a 6xHis tag. The promoter driving the expression is the tightly-regulated, methanol-induced *AOX1* promoter. This vector does not contain a yeast origin of replication and thus transformants can only be isolated if recombination occurs between the plasmid and the *Pichia* genome. For cloning, ORF P1192R was PCR-amplified skipping the endogenous stop codon, hydrolyzed with *Kpn*I and *Not*I and cloned in this plasmid, previously hydrolyzed with the same restriction enzymes, in-frame with the existent 5' and 3' features. This originated plasmid pPICZαC-P1192R.

To express P1192R in *Saccharomyces cerevisiae*, the complete ORF P1192R was PCR-amplified with forward and reverse primers containing *Kpn*I and *Not*I sites, respectively. The PCR fragment was hydrolyzed and cloned in the corresponding sites of the yeast expression vector pYES2 (kindly provided by Dr. Luísa Cyrne, University of Lisbon, Portugal), which is a 2µm-based vector, originating plasmid pYES2-P1192R. Driving the expression in pYES2 is the *S. cerevisiae GAL1* promoter. Transcription from this promoter is strongly repressed in the presence of glucose, but in the presence of galactose and absence of glucose expression can be up-regulated up to 2000-fold (Maya, Quintero, de la Cruz Muñoz-Centeno, & Chávez, 2008). To obtain the catalytic mutant in pYES2-P1192R site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies), according to the manufacturer's instructions, including the primer design. This implied the mutation of an adenine at position 2399 to a

thymine, originating a tyrosine to phenylalanine change in the respective amino acid sequence. To facilitate downstream purification of recombinant pP1192R by affinity chromatography, it was necessary to add a 6xHis tag to the P1192R ORF. This was done by using plasmid pPICZαC-P1192R as a starting template. Using the same forward primer used for cloning in pPICZαC in combination with a new specific reverse primer, with a *Not*I site, the P1192R-6xHis ORF was PCR amplified and cloned in vector pYES2, originating plasmid pYES2-P1192R_6xHis. For addition of a GST tag to P1192R, the ORF coding for GST was PCR-amplified from plasmid pGEX-6P-1, with the inclusion of a stop codon and *Not*I sites up- and downstream of the ORF. The polyhistidine tag present in pYES2-P1192R_6xHis was removed by hydrolysis with *Not*I and the GST fragment, previously hydrolyzed with the same restriction enzyme, was inserted in its place. This originated plasmid pYES2-P1192R_GST.

To obtain a positive control for the complementation of the temperature-sensitive yeast strains, the yeast ORF coding for the wild-type endogenous type II topoisomerase (*TOP2*) was cloned in the pYES2 vector. For this, ORF *TOP2* was PCR-amplified (skipping the endogenous stop codon) from *S. cerevisiae* genomic DNA using primers with built-in *Bam*HI and *Xho*I sites and cloned in the pET24a vector, in-frame with the present 6xHis tag, using these two enzymes. The same forward primer in combination with a new reverse primer, with a *Not*I site, were used for PCR amplification of the *TOP2*-6xHis ORF, which was cloned in pYES2, originating the plasmid pYES2-TOP2.

For the construction of the P1192R-TOP2C-terminal fusion protein, plasmid pYES2-P1192R_6xHis was used as a starting template. Hydrolysis of this plasmid with *Not*I resulted in the excision of the 6xHis tag, since it was located between *Not*I sites. Then, the *TOP2* C-terminal region (nucleotides 3543-4287) was PCR-amplified using primers containing *Not*I sites and cloned in frame with P1192R, thereby substituting the polyhistidine tag and originating plasmid pYES2-P1192R-TOP2Cterm. The catalytic mutant in pYES2-P1192R-TOP2Cterm was obtained as described above for pYES2-P1192R.

Cloning of the *TOP2* C-terminal region in pYES2, to use as a negative control, was performed by obtaining the respective fragment as described above and cloning it in a *Not*I-hydrolyzed pYES2 vector, originating plasmid pYES2-TOP2Cterm.

To analyze the cellular localization of pP1192R and pP1192R-Top2Cterm, the GFP ORF from pIC113 was amplified by PCR using primers containing *Kpn*I sites, hydrolyzed with this restriction enzyme and cloned in *Kpn*I-hydrolyzed plasmids pYES2, pYES2-P1192R or pYES2-P1192R-TOP2Cterm, respectively.

### 4.2.4 Bacterial strains and transformation

Maintenance and transformation of bacterial strains were performed essentially as indicated on Chapter 2. Vector pPICZαC contains, as a selection marker, the *Sh ble* gene conferring resistance to the antibiotic zeocin (Drocourt, Calmels, Reynes, Baron, & Tiraby, 1990; Baron, Reynes, Stassi, & Tiraby, 1992). This drug intercalates DNA molecules and cleaves them, inducing cell death both in prokaryotes and eukaryotes. Therefore, this selection marker can be used for selection of transformants in *E. coli* and in *P. pastoris* cells. If used for selection in *E. coli* the salt concentration of the LB medium must remain low (below 90 mM) and therefore the medium recipe had to be adapted accordingly. Selection with zeocin in low salt LB was performed by supplementing the medium with 25 µg/ml of antibiotic.

For protein expression assays, *E. coli* expression strains BL21(DE3)pLysS, Rosetta(DE3)pLysS, C41(DE3) and C43(DE3) were used. Strains containing pLysS were kept as described previously but the medium was supplemented with chloramphenicol (at 34 µg/ml) for maintenance of the plasmid.

### 4.2.5 Bacterial cultures for protein expression

All the cultures for protein expression in *E. coli* were prepared using LB as culture medium, supplemented with 30 µg/ml kanamycin for selection of pET24a-based plasmids. To facilitate the procedure, the bacterial strain was first transformed with the appropriate plasmid(s) and the transformants were plated on LB medium supplemented with the appropriate antibiotic(s). From these plates, colonies were selected to inoculate starter cultures, usually of 5 - 10 ml of LB containing the appropriate antibiotic(s), which were incubated overnight (around 16 hours) at 37 ºC, at 180 - 200 rpm.

### 4.2.6 Expression and purification of recombinant pP1192R from *Escherichia coli*

The initial procedure was as follows: recombinant pP1192R was expressed in *E. coli* strain BL21(DE3)pLysS, after 3 hours of induction with 1 mM IPTG (Sigma-Aldrich) and incubation at 37 ºC. Cells from 200 ml of culture were collected by centrifugation at 8000 *g* for 5 minutes, 4 ºC, and resuspended in 5 ml of lysis buffer [20 mM Tris-HCl pH 8, 500 mM NaCl, 10 mM imidazole, 2% (v/v) fos-choline-12 (Anatrace)]. Cells were lysed by performing three freeze/thaw cycles and passing the suspension 4 - 5 times through a 26G x 1/2'' 0.45 mm x 12 mm needle. After 10 minutes of centrifugation at 13200 rpm, to separate cellular debris, the supernatant was collected and, while kept on ice, subjected twice to sonication in a Sonopuls HD2070 ultrasonic homogenizer (Bandelin) for 10 minutes with 50% cycle and 50% power. After a centrifugation at 3000 *g* for 30 minutes, 4 ºC, the

supernatant was collected and 600 µl fractions were used to load a His SpinTrap column (GE Healthcare). Purification was performed as indicated by the manufacturer. Two washing buffers were used, similar to the lysis buffer but containing either 50 mM or 100 mM imidazole. Protein was eluted from the column with elution buffer, similar to the lysis buffer but containing 500 mM imidazole. Protein aliquots were stored at -80ºC until further use.

Modifications of the initial procedure are described along the Results section.

### 4.2.7 Maintenance of *Pichia pastoris*

The *P. pastoris* strain used in this work was X33, kindly provided by Dr. Rodrigo Cunha, Pelotas, Brazil. It is a wild-type strain, able to grow on complete or minimal media and useful for selection with zeocin. Unless stated otherwise, the wild-type strain or transformants derived from it were grown at 28 - 30 ºC. When untransformed, this strain was kept and grown on YPD medium [2% (w/v) Bacto Peptone (BD Biosciences), 1% (w/v) yeast extract (Biokar Diagnostics), 2% (w/v) glucose (Sigma-Aldrich)], also containing 2% (w/v) agar (Sigma-Aldrich) when solid. Following transformation, the strain was grown on solid YPDS medium [YPD supplemented with 1 M sorbitol (Sigma-Aldrich)] containing 100 µg/ml zeocin (Invivogen). Minimal Methanol (MM) medium [1.34% (w/v) yeast nitrogen base (YNB, Sigma-Aldrich), 0.5% (v/v) methanol (Merck), 4 x $10^{-5}$% (w/v) biotin (Sigma-Aldrich)] was used for selection of $Mut^+$ transformants. After selection, transformants were kept on solid YPD medium and grown on liquid BMGY medium [1% (w/v) yeast extract, 2% (w/v) Bacto Peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x $10^{-5}$% (w/v) biotin, 1% (v/v) glycerol (Sigma-Aldrich)].

### 4.2.8 Transformation and selection of *Pichia pas*toris

*P. pastoris* cells were transformed by electroporation, following the protocol recommended by the manufacturer (EasySelect *Pichia* Expression kit, Invitrogen). For this, a 5 ml YPD pre-culture was inoculated and grown at 30 ºC overnight, after which 0.1 - 0.5 ml of this culture were used to inoculate a 500 ml culture. The latter was also grown overnight, to an $OD_{600}$ of 1.3 - 1.5. Cells were then collected and centrifuged at 2000 *g* for 5 minutes, at 4 ºC. After discarding the supernatant, the pellet was resuspended in 500 ml of ice-cold sterile $H_2O$ and centrifuged as before. The supernatant was again discarded, the pellet resuspended in 250 ml of ice-cold sterile $H_2O$ and centrifuged as before. The supernatant was again discarded, the pellet was resuspended in 20 ml of ice-cold 1 M sorbitol and centrifuged as before. The supernatant was discarded and the pellet was resuspended in 1 ml of ice-cold 1 M sorbitol. From this cell suspension, 80 µl were used per transformation and mixed with

5 - 10 µg of *Age*I- or *Pme*I-linearized plasmid DNA, and the cell-DNA mixture was transferred to an ice-cold 0.2 cm electroporation cuvette (Sigma-Aldrich). Cuvettes were kept on ice for 5 minutes and cells were pulsed in a Gene Pulser II Electroporation System (Bio-Rad) with 1.5 kV, 200 Ω, 25 µF, for a pulse time of about 5 milliseconds. This was immediately followed by addition of 1 ml of ice-cold 1 M sorbitol solution to each cuvette, and the entire contents were transferred to polypropylene 15 ml tubes and incubated at 30 ºC, without shaking, for 1 - 2 hours. After this time, 200 µl of each cell suspension were used to plate cells on YPDS solid medium containing 100 µg/ml of zeocin. Plates were then incubated at 30 ºC for 5 - 6 days, until the appearance of colonies. When selection of multi-copy recombinants was desired, concentration of zeocin in the medium was increased to 250, 500 or 1000 µg/ml. Once zeocin-resistant transformants are obtained, it is not necessary to maintain the recombinant *P. pastoris* clone in medium containing zeocin for expression studies. Thus, zeocin is only required for initial screening and selection of recombinant clones.

For confirmation of the Mut phenotype, colonies resulting from the transformation procedure were picked and streaked sequentially on solid MM and YPDS media. This order was mandatory because traces of glucose from the YPDS may inhibit the methanotrophic metabolism in MM medium. Plates were incubated for two days at 28 - 30 ºC.

### 4.2.9 DNA extraction from yeast cells

Genomic DNA was extracted from *P. pastoris* or *S. cerevisiae* cells using an adaptation of the Bust n' Grab method (Harju, Fedosyuk, & Peterson, 2004). Cells from 1.5 ml of an overnight-grown YPD culture were collected by centrifugation for 5 minutes, 13200 rpm, and the supernatant was discarded. The pellet was resuspended in 200 µl of Harju-buffer [10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 2% (w/v) Triton X-100 (Sigma-Aldrich), 1% (w/v) SDS, 1 mM ethylenediaminetetraacetic acid (EDTA, VWR BDH Prolabo)], the tube was incubated at -70 ºC for 2 minutes and thawed at 95 ºC for 1 minute, and these last two steps were repeated twice. The tube was then vortexed for 10 seconds and 200 µl of chloroform were added, followed by vortexing for 5 - 10 minutes. Following centrifugation for 15 - 20 minutes, 13200 rpm, the upper aqueous phase was collected to a clean tube and to it 400 µl of ice-cold absolute ethanol were added, and the tube contents were mixed by repeated inversion. After one hour of incubation at -70 ºC, the tube was centrifuged for 15 minutes, 13200 rpm, the supernatant was discarded and the pellet washed with 750 µl of 70% (v/v) ethanol. After drying, the pellet was resuspended in 50 µl of water and DNA concentration was measured.

### 4.2.10 Protein extraction from yeast cells for immunoblotting

The method of total protein extraction from yeast cells for immunoblotting purposes was adapted from Kushnirov (2000). Cells from 1 ml of culture were collected by centrifugation at 13200 rpm for 2 minutes and resuspended in a 0.1 M NaOH solution. After 30 minutes of incubation, cells were centrifuged at 7000 rpm for 5 minutes and resuspended in an appropriate volume of protein loading buffer for a final concentration of 1X. After boiling for 5 minutes and centrifuging at 13200 rpm for 1 minute, the desired volume of extract was loaded on an SDS-PAGE gel.

### 4.2.11 Confirmation of *Pichia* integrants by PCR

Genomic DNA extracted from *P. pastoris* cultures was screened for integration of desired constructs by PCR using specific primers. Since proofreading wasn't required, DreamTaq DNA Polymerase was used. Per reaction, 1 µl of genomic DNA was used, by adding it to a PCR tube containing 24 µl of colony PCR mixture (1X Dream Taq Buffer Green, 200 µM dNTPs, 0.5 µM of each primer, 1.25 U of DreamTaq DNA Polymerase). The tubes were placed on a Mastercycler gradient thermal cycler and subjected to the following conditions: tubes were heated to 95 ºC for 2 minutes followed by 25 cycles of denaturation at 95 ºC for 1 minute, annealing 55 ºC for 1 minute and extension at 72 ºC for 4 minutes. A final extension step of 10 minutes at 72 ºC concluded the PCR cycling conditions.

### 4.2.12 Expression of P1192R in *Pichia pastoris*

Heterologous expression using *P. pastoris* transformants confirmed to have integration of the desired construct was done by essentially following the protocol recommended by the manufacturer (EasySelect *Pichia* Expression kit, Invitrogen). A single colony was used to inoculate 25 ml of BMGY in a 500 ml flask. After overnight incubation at 30 ºC, 250 rpm, cells were collected by centrifugation at 2000 *g* for 5 minutes, the supernatant was discarded and the cell pellet was resuspended in 25 ml of BMMY [1% (w/v) yeast extract, 2% (w/v) Bacto Peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x $10^{-5}$% biotin, 0.5% (v/v) methanol]. From this suspension, the required volume was used to inoculate 100 ml of BMMY at $OD_{600}$ = 0.1, in a 1 liter flask with a cotton plug to facilitate oxygenation. The culture was incubated at 30 ºC, 250 rpm, for up to 96 hours, with methanol being added every 24 hours for a final concentration of 0.5% (v/v), to maintain induction. At several time points during induction (0, 6, 12, 24, 36, 48, 60, 72, 84, and 96 hours), samples of 1 ml of culture were collected to 1.5 ml tubes, centrifuged for 2 minutes at 13200 rpm, the

supernatant was collected to a new tube and both the supernatants and the cell pellets were stored at -80 ºC until processing.

### 4.2.13 Maintenance of *Saccharomyces cerevisiae*

*S. cerevisiae* strains used in this work were JCW26 (*MATα leu2 his3-200 trp1-63 ura3-52 top2-4*), SD117 (*MATα ade2 leu2 his7 trp1 ura3-52 top2-1*) (Wasserman, Austin, Fisher, & Wang, 1993), kindly provided by Dr. Caroline Austin, Newcastle University, United Kingdom, and BY4741 (*MATa met15Δ0 leu2Δ0 his3Δ1 ura3Δ0*) (Brachmann *et al.*, 1998), kindly provided by Dr. Luísa Cyrne, University of Lisbon, Portugal. Strains JCW26 and SD117, being thermo-sensitive, were always grown at 25 ºC. The wild-type strain BY4741 was grown at 28 - 30 ºC, unless stated otherwise. When untransformed, all the strains were kept and grown on YPD medium, also containing 2% (w/v) agar when solid. When transformed, and unless stated otherwise, the strains were grown on solid or in liquid casamino acid (CAA) medium without uracil [0.6% (w/v) casamino acids (BD Diagnostics), 0.685% (w/v) YNB, 0.0025% (w/v) adenine (Sigma-Aldrich), 0.005% (w/v) L-tryptophan (Sigma-Aldrich)] and with the addition of 2% (w/v) glucose. Glucose was substituted by raffinose (Sigma-Aldrich), also at 2% (w/v), when an alternative, non-repressing carbon source was required for cell growth prior to induction with galactose.

### 4.2.14 Transformation of *Saccharomyces cerevisiae*

Transformation of *S. cerevisiae* cells was done by following the method of Ito *et al.* (1983), with modifications. For this, a 100 ml YPD culture was inoculated at $OD_{600}$ = 0.1, using the appropriate volume of a pre-culture grown at 25 ºC overnight, and incubated at 25 ºC, 180 - 200 rpm, until the culture reached an $OD_{600}$ of about 0.4. At this point 10 ml fractions of the culture were collected into 15 ml polypropylene tubes and centrifuged at 3000 *g* for 5 minutes. After discarding the supernatants, each cell pellet was resuspended in 10 ml of sterile $H_2O$ and tubes were centrifuged again at 3000 *g* for 5 minutes. After discarding the supernatants, each cell pellet was resuspended in 1 ml of sterile $H_2O$ and transferred to a sterile 1.5 ml microtube. After centrifugation for 5 minutes at 4000 rpm, the supernatants were discarded and each cell pellet was resuspended in 1 ml of a LiAc/TE solution [10 mM lithium acetate (Sigma-Aldrich), 10 mM Tris pH 8 and 1 mM EDTA (VWR BDH Prolabo)]. After centrifugation for 5 minutes at 4000 rpm, the supernatants were discarded and each cell pellet was resuspended in 50 µl of LiAc/TE solution. Each 50 µl suspension was used for a transformation reaction, by adding to it 50 µg of single-stranded salmon sperm DNA (Sigma-Aldrich), 1 - 2 µg of purified plasmid and 300 µl of a LiAc/TE

solution supplemented with 40% (w/v) PEG 3350 (Sigma-Aldrich). After vortexing briefly, tubes were incubated for 30 minutes at 25 ºC, 180 - 200 rpm, followed by a heat-shock for 15 minutes at 42 ºC. After cooling on ice for 2 minutes, tubes were centrifuged at 4000 rpm for 5 minutes, the supernatants were discarded and each cell pellet was resuspended in 150 µl of sterile $H_2O$. The entire volume was then spread on a CAA without uracil plate (one plate for each transformation reaction), and plates were incubated for 5 days at 25 ºC until colonies were visible.

### 4.2.15 Production of polyclonal anti-pP1192R serum using purified protein expressed in *Saccharomyces cerevisiae*

*S. cerevisiae* strain BY4741 harboring the recombinant pYES2-P1192R_6xHis plasmid was grown in CAA medium without uracil to an $OD_{600}$ of 1 and induction was carried out by addition of galactose to a final concentration of 6% (v/v) for 24 hours, at 25 ºC. Cells were then harvested by centrifugation at 6000 *g* for 5 minutes, washed with sterile $H_2O$ and the pellet was resuspended in buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 1 mM PMSF). About half the volume of acid-washed glass beads (G8772, Sigma-Aldrich) was added and a single 30-minute cycle of vortexing was performed, while adding 0.1 mM PSMF every 5 minutes. The supernatant was collected, the beads washed with buffer A and both supernatants were pooled and centrifuged for 30 minutes at 15000 *g*, followed by filtration using a 0.45 µm filter. The extract was then mixed with 0.2 volumes of buffer B (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 250 mM imidazole, 10 mM beta-mercaptoethanol) and the mixture was incubated with Ni Sepharose 6 Fast Flow slurry (GE Healthcare) for about 3 hours. The mixture was loaded onto a PD-10 column (GE Healthcare), washed sequentially with 10 bead volumes of a combination of buffers A and B (8%, 20% and 30% buffer B), and elution was performed using 5 bead volumes of buffer B. Fractions were collected in low-binding tubes, analyzed by SDS-PAGE and stored at -80 ºC.

For inoculation, the protein was thawed, quantified and 10 µg and 25 µg were mixed with an equal volume of Freund's incomplete adjuvant (and PBS, when necessary) for a total volume of 500 µl. The emulsified samples were injected intraperitoneally into CD1 mice. The mice were given booster injections after two and four weeks, sacrificed two weeks later and their blood was collected and the serum separated.

### 4.2.16 Complementation assays

To carry out complementation assays in solid medium, *S. cerevisiae* cells from CAA without uracil with 2% (w/v) raffinose cultures at stationary phase were collected by

centrifugation and resuspended in previously sterilized disaggregation buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA), and the cell density was adjusted to an $OD_{600}$ of 3. From this suspension, serial three-fold dilutions were prepared, in disaggregation buffer, and 3 µl of each dilution were spotted on freshly prepared solid YP medium supplemented with either 2% (w/v) glucose or 2% (w/v) galactose (Sigma-Aldrich) as the carbon source. The plates were incubated for 5 days at 25 ºC or 35 ºC.

For the complementation assays in liquid medium, *S. cerevisiae* cells from CAA without uracil cultures with 2% (w/v) raffinose cultures at stationary phase were used to inoculate new CAA without uracil with 2% (w/v) galactose cultures, at an $OD_{600}$ of 0.1. These cultures were split in half, with one half being incubated for 24 hours at 25 ºC while the other was incubated for the same time period at 35 ºC. Samples for the cultures were collected at 6, 12 and 24 hours post-induction for determination of their growth curves.

### 4.2.17 Fluorescence microscopy of *Saccharomyces cerevisiae*

*S. cerevisiae* cells in stationary phase from CAA without uracil with 2% (w/v) raffinose cultures were used to inoculate new CAA without uracil cultures with 2% (w/v) galactose cultures, at an $OD_{600}$ of 0.5. After 16 hours of incubation at 25 ºC, 450 µl of each were collected and to each 50 µl of formol (36.5% formaldehyde, 10% methanol) were added. After 10 minutes of incubation at room temperature cells from these formaldehyde-containing suspensions were collected by centrifugation and the pellets were resuspended in 1 ml PBS containing 10% (v/v) formol. After 6 hours of incubation at room temperature cells were collected by centrifugation, washed three times with PBS and incubated with 1 µg/ml DAPI for 10 minutes. Equal volumes of each suspension and of Fluoromount-G were mixed and 5 µl of that mixture were used to prepare samples for visualization.

### 4.2.18 Preparation of soluble protein extracts from *Saccharomyces cerevisiae*

Cells from *S. cerevisiae* strain JCW26 containing plasmid pYES, pYES2-P1192R, pYES2-P1192R Y800F, pYES2-P1192R-TOP2Cterm or pYES2-TOP2, were grown in CAA medium without uracil containing 2% (w/v) raffinose, to an $OD_{600}$ of 0.4 and induction was carried out by addition of galactose to a final concentration of 2% (w/v) for 22 hours, at 25 ºC. Cells were then harvested by centrifugation and each cell pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 1 mM EDTA, 10 mM $(NH_4)_2SO_4$, 0.1% (w/v) Triton X-100, 1 mM DTT, 1 mM PSMF), with the volume of lysis buffer depending on the cell culture density at the time of collection, and about half the volume of acid-washed glass beads was added. After a single 30-minute cycle of vortexing,

with the addition of 0.1 mM PSMF every 10 minutes, each suspension was subjected twice to 1 minute of sonication. From each suspension 1 ml was collected and centrifuged at 16000 *g* for 10 minutes, 4 ºC, and the supernatants were saved. For enzymatic assays different amounts of supernatant were used.

### 4.2.19 Immunoblotting

Loading and running SDS-PAGE gels, as well as transfer of proteins to nitrocellulose membranes, were performed as described on Chapter 3. The molecular weight marker used was either NZYColour protein marker II (NZYTech), henceforth designated as CPM, or Precision Plus Protein Standards (Bio-Rad), from now on mentioned as PPPS.

Immunoblots were blocked for 2 - 3 hours with either TBS or PBS containing 5% (w/v) non-fat milk and 0.1% (w/v) Tween 20 before being probed with primary and secondary antibodies or anti-sera. When using the anti-pP1192R serum, 5% (w/v) of commercial yeast extract was sometimes added to the blocking buffer. Finally, membranes were developed using chemiluminescent reagents SuperSignal West Pico or SuperSignal West Femto (ThermoScientific), according to the manufacturer`s instructions. The primary antibodies used were a mouse polyclonal antibody raised against the recombinant full-length pP1192R produced in *S. cerevisiae* (described in detail on this Chapter) and a commercial anti-6xHis antibody (27-4710-01, GE Healthcare), at 1:1000 and 1:2000, respectively. The conjugated secondary antibody was an anti-mouse-HRP conjugated (A2554, Sigma-Aldrich), at 1:50000. All dilutions of primary and secondary antibodies were performed in blocking solution.

### 4.2.20 Expression and purification of functional recombinant pP1192R from *Saccharomyces cerevisiae*

*S. cerevisiae* strain BY4741 harboring the recombinant pYES2-P1192R_6xHis plasmid was grown overnight in CAA medium without uracil and containing 2% (w/v) raffinose. This pre-culture was used to inoculate a 500 ml culture of CAA medium without uracil containing 2% galactose, at an $OD_{600}$ of 1 and induction was carried out for 23 hours, at 28 - 30 ºC. Cells were then harvested by centrifugation at 8000 *g* for 5 minutes, washed with sterile $H_2O$ and the pellet was resuspended in 10 ml of lysis buffer [50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 0.1 mM PMSF]. From this point, all of the procedure was performed either in the cold room (4 - 6 ºC) or under refrigerated conditions (centrifugations). About one volume of acid-washed glass beads was added to the cell suspension and six cycles of 5 minutes of vortexing were performed, with each cycle being followed by 5 minutes of incubation on ice, with 0.1 mM PMSF being added every

10 minutes. After a first centrifugation for 10 minutes at 3500 *g*, 4 ºC, to clear the extract from cell debris and beads, the supernatant was collected and centrifuged again for 30 minutes at 15000 *g*, 4 ºC, followed by filtration using a 0.45 µm filter. The extract was then mixed with 0.2 volumes of buffer C 5x [50 mM sodium phosphate buffer pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 100 mM imidazole] and the mixture was incubated with pre-equilibrated Ni Sepharose 6 Fast Flow slurry (GE Healthcare) for about 3.5 hours. The mixture was loaded onto a PD-10 column (GE Healthcare), washed sequentially with 16 bead volumes of a combination of buffers A [10 mM sodium phosphate buffer pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 10 mM beta-mercaptoethanol] and B (similar to A but containing 250 mM imidazole) for final proportions of 10%, 20% and 30% of buffer B (corresponding to 25, 50 and 75 mM imidazole, respectively), and elution was performed using 3.5 bead volumes of buffer B. Fractions were collected in low-binding tubes and after analysis by SDS-PAGE and topoisomerase activity assays, fractions found to contain topoII activity were pooled, quantified in triplicate using the Bradford protein assay (Bio-Rad), aliquoted and stored at -80 ºC.

### 4.2.21 Assays of type II DNA topoisomerase activity

Type II DNA topoisomerase decatenation activity was tested by using kinetoplast DNA (kDNA) from *Crithidia fasciculata* (TopoGEN). Reactions were carried out for up to 1 hour at 37 ºC in a total of 20 µl of reaction buffer containing 150 ng of kDNA, with or without 2 mM ATP (Sigma-Aldrich). Reactions were stopped by addition of 4 µl of stop solution [5% (v/v) sarcosyl, 0.003% (w/v) bromophenol blue, 25% (v/v) glycerol] and electrophoresed in a 0.8 - 1.0% (w/v) agarose gel, with TAE buffer (40 mM Tris-Acetate, 1 mM EDTA), for 1.5 - 2 hours at 90 V or 16 hours at 20 V.

Relaxation activity was tested by using either negatively supercoiled pBR322 plasmid or negatively supercoiled pRYG plasmid (both from TopoGEN). Reactions were carried out for up to 1 hour at 37 ºC in a total of 20 µl of reaction buffer containing 5 nM of supercoiled plasmid, with or without 2 mM ATP (Sigma-Aldrich). Reactions were stopped by addition of 4 µl of stop solution and electrophoresed in a 1.0% (w/v) agarose gel, with TAE buffer, for 16 hours at 20 - 25 V.

Supercoiling and/or catenation activity were tested using relaxed pBR322 plasmid (TopoGEN). Reactions were carried out for 30 minutes at 37 ºC in a total of 20 µl of reaction buffer containing 5 nM of relaxed plasmid and 2 mM ATP. Reactions were stopped by addition of 4 µl of stop solution and electrophoresed in a 1.0% (w/v) agarose gel, with TAE buffer, for 16 hours at 20 - 25 V.

Until the purification of recombinant pP1192R expressed in *Saccharomyces cerevisiae*, the reaction buffer used for decatenation or relaxation assays was the following: 50 mM Tris-HCl pH 7.5, 20 mM NaCl, 20 mM $MgCl_2$, 0.5 mM EDTA, 1 mM ATP, 0.5 mM DTT. After the above mentioned purification, and for the characterization of the optimal *in vitro* conditions for the activity of pP1192R, the reaction buffer used was: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM ATP and 1 mM DTT. Once the optimal working conditions were defined, decatenation assays to study the effects of topoisomerase inhibitors were performed using the following reaction buffer: 50mM Tris-HCl pH 7.5, 75 mM NaCl, 6 mM $MgCl_2$, 2 mM ATP and 1 mM DTT.

The effects of pH, ionic strength, divalent ion concentration, ATP concentration and temperature were measured using relaxation assays, with DNA relaxation being determined from the loss of supercoiled DNA and accumulation of relaxed DNA. Each variable was changed independently of the others. Of note, the pH of the buffers used was set at 25 ºC and no attempt was made to correct for its variation due to temperature effects during the assays. Likewise, in the experiments varying temperature, no attempt was made to control the pH variation at each different temperature.

Topoisomerase poisons and inhibitors tested were: ICRF-193 (I4659), etoposide/VP-16 (E1383), *m*-AMSA (A9809), coumermycin A1 (C9270), genistein (G6649), naringenin (N5893), doxorubicin (D1515), topotecan (T2705), nalidixic acid (N8878), ciprofloxacin (17850), norfloxacin (N9890), enrofloxacin (17849), and gatifloxacin (32345), all purchased from Sigma-Aldrich. Unless stated otherwise, these were tested at the following concentrations: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512 and 1000 µM.

All reactions were prepared on ice, with the enzyme or total extract always being the final component to be added. Reactions were considered to start once incubation at the desired temperature was initiated.

## 4.3 Results

### 4.3.1 Heterologous expression of P1192R in *Escherichia coli*

We previously observed (Chapter 3) that P1192R codes for a protein expressed during viral infection. However, it remained to be demonstrated that pP1192R is indeed a functional type II DNA topoisomerase. As described in Chapter 2, we cloned ORF P1192R in the pET24a expression vector, in frame with the existent 6xHis tag. Here, we used this clone to attempt heterologous expression of this ORF in *E. coli*, using expression strains BL21(DE3)pLysS and Rosetta(DE3)pLysS. The latter strain is based on the former but contains, in plasmid pLysS, tRNA genes coding for codons rare in *E. coli* but common in

eukaryotic organisms (Gopal & Kumar, 2013). In small-scale preliminary expression attempts, the levels of expressed protein using BL21(DE3)pLysS or Rosetta(DE3)pLysS were very similar. This indicated that codon usage was not problematic for expression of P1192R and we chose to continue using the BL21(DE3)pLysS strain. Several conditions (IPTG concentration, induction period, temperature of incubation) were tested using this strain and the optimal conditions were found to be a 3 hour incubation period, at 37ºC, with a concentration of 1mM IPTG. Next, we tested extraction of the recombinant protein under native conditions, since our main goal was to obtain functional pP1192R to demonstrate its activity as a type II DNA topoisomerase. However, initial extraction attempts were unproductive and the protein was found to be insoluble, probably accumulating in inclusion bodies. Because extraction of proteins from inclusion bodies usually requires harsh, denaturing conditions (Singh, Upadhyay, Upadhyay, Singh, & Panda, 2015), that disrupt protein structure and may compromise activity, before doing so we decided to try two other *E. coli* strains, C41(DE3) and C43(DE3), for expression of pP1192R. Strain C41(DE3) is derived from BL21(DE3) and contains at least one mutation that prevents cell death associated with expression of recombinant toxic proteins (Miroux & Walker, 1996). Strain C43(DE3) is derived from C41(DE3) by selection for resistance to a different toxic protein, and is therefore able to express a different set of toxic proteins (Miroux & Walker, 1996). These strains have also been described to express recombinant proteins while avoiding their accumulation in inclusion bodies (Arechaga *et al.*, 2000). In contrast with the BL21(DE3)pLysS strain which we were using, strains C41(DE3) and C43(DE3) did not contain the pLysS plasmid and low-level expression could occur in the absence of the inducer. Expression of P1192R was tested in C41(DE3) and C43(DE3), using the conditions found to be optimal for BL21(DE3)pLysS, but we did not detect expression of recombinant pP1192R. We thus returned to BL21(DE3)pLysS and attempted extraction of pP1192R under denaturing conditions, either through the use of detergents or urea. Detergents used in this phase were Tween 20, Triton X-100, sodium lauroyl-sarcosinate (sarcosyl), NP-40, CHAPS, N-dodecylphosphocholine (fos-choline-12), CYMAL-5, N-decyl-β-D-maltopyranoside, N-octyl-β-D-glucopyranoside and N-dodecyl-β-D-maltopyranoside, all at a maximum concentration of 2% in the lysis buffer. Of these, only sarcosyl and fos-choline-12 were efficient in promoting solubilization of recombinant pP1192R. While the latter was sufficient when present at 0.5%, the former was required at 2% in the lysis buffer to have a significant effect on the solubility of recombinant pP1192R (Figure 33).

However, we could not find any clear information on the compatibility of the nickel matrix with either detergent. For fos-choline-16, data from QIAGEN mentioned 0.05% of detergent,

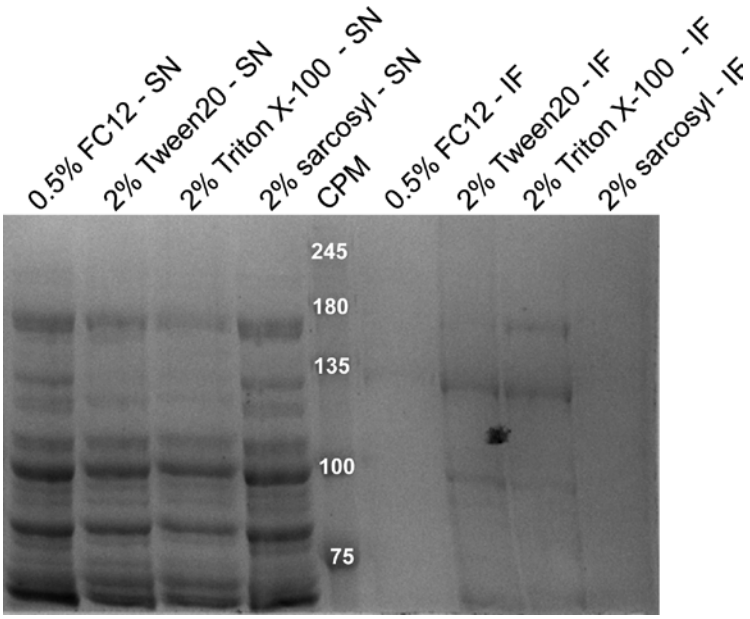


**Figure 33.** Solubilization of recombinant pP1192R through the use of detergents.
Soluble (SN) or insoluble (IF) fractions of total protein extracts from *E. coli* cells expressing pP1192R were analyzed in a 7.5% SDS-PAGE gel, followed by Coomassie staining. FC12 is an abbreviation for fos-choline-12. CPM is the molecular weight marker.

which was ten times lower than our working 0.5% for fos-choline-12. Also, while SDS, which is similar to sarcosyl, is not recommended for use with the nickel matrix, the same data from QIAGEN mentioned 0.3% SDS having been used successfully. To assess if we would be able to purify the protein using the defined optimal conditions for expression and extraction, we performed rapid small-scale purification using microcentrifuge His SpinTrap columns. This was performed using fos-choline-12, since it is a zwitterionic detergent, as opposed to sarcosyl which is anionic, and therefore we hoped that fos-choline-12 wouldn't (partially or totally) disrupt the structure of pP1192R and compromise its activity. An additional plus was that we could have fos-choline-12 present in the buffers at a lower concentration than sarcosyl. We confirmed that the protein corresponding to a band with the expected molecular weight, which appears upon addition of IPTG to the culture medium and becomes soluble in the presence of sarcosyl or fos-choline-12 in the lysis buffer, does indeed bind the Ni Sepharose matrix of the column, being eluted only after addition of a high imidazole concentration (500mM) (Figure 34A). In a Western blot using an antibody recognizing the 6xHis tag we also detected a band of identical molecular weight (Figure 34B), supporting its identification as recombinant pP1192R.

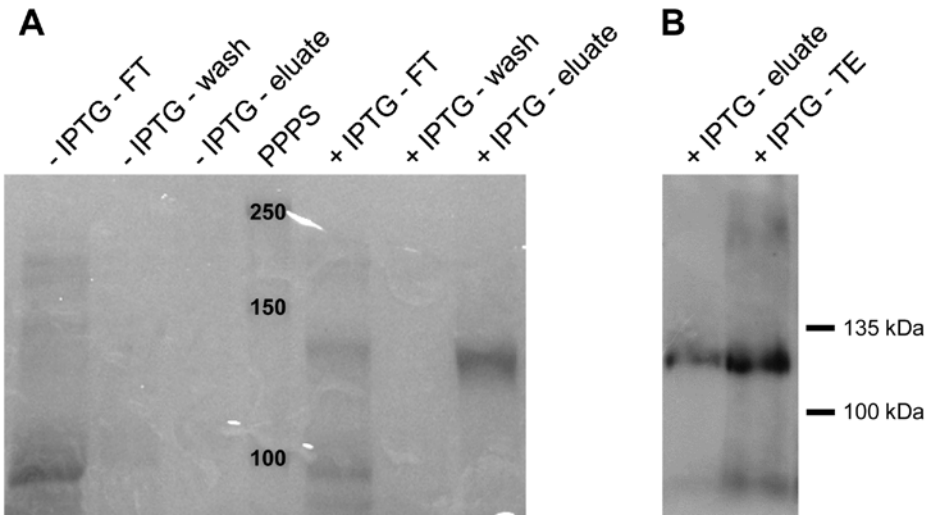


**Figure 34.** Solubilization of recombinant pP1192R through the use of fos-choline-12.
A) Fos-choline-12-containing soluble fractions from total protein extracts of non-induced (- IPTG) and induced (+ IPTG) BL21(DE3)pLysS cells transformed with pET24a-P1192R_6xHis were loaded on His SpinTrap columns and the respective flowthrough (FT), wash and eluate fractions were analyzed in a 7.5% SDS-PAGE gel, followed by Coomassie staining. PPPS is the molecular weight marker.
B) The eluate fraction from the His SpinTrap purification, as in A), as well as a total extract of BL21(DE3)pLysS cells expressing recombinant pP1192R were analyzed by Western blot using a commercial anti-6xHis antibody.

### 4.3.2 Purification of pP1192R expressed in *E. coli* and *in vitro* activity assays

We then increased the scale of expression and attempted to purify the expressed protein from the induced cells using 0.5% fos-choline-12. When this was followed by on-column refolding attempts, by gradually decreasing the concentration of detergent during the washes and using elution buffer without fos-choline-12, the protein was not eluted upon increase of imidazole concentration. When elution in the presence of detergent was followed by dialysis against PBS, no activity was observed in *in vitro* relaxation assays using supercoiled pBR322.

Purification of pP1192R after lysis with 2% sarkosyl was also attempted. Though the protein bound efficiently to the nickel resin and was successfully eluted from it, it showed no signs of being functional in a similar relaxation assay.

We then tried extracting and purifying the protein using denaturing conditions with urea, which was easily dialyzed. Yet, protein obtained through this method was also inactive in *in vitro* relaxation assays.

### 4.3.3 Cloning of P1192R, transformation and selection of *Pichia pastoris*

Given the unsuccessful attempts at obtaining functional recombinant pP1192R expressed in *E. coli* cells, a change of expression system was mandatory. As such, *Picha pastoris*

presented itself has a good alternative to *E. coli* for several reasons. It's an eukaryotic organism, and since viral pP1192R is expressed in a eukaryotic-cell context (either mammalian or tick cells), it's more likely for pP1192R to be correctly expressed, processed, folded and assembled into functional molecules in a eukaryotic expression system. In addition, *P. pastoris* is a single-celled organism and relatively fast and inexpensive to grow and manipulate in culture, and most of the genetic engineering tools and knowledge which have been generated for *S. cerevisiae* can be applied. Efficient secretion of recombinant proteins and high levels of heterologous expression using the *AOX1* promoter were also decisive factors leading to the choice of this system.

To produce recombinant pP1192R in *P. pastoris* it was first necessary to clone its corresponding ORF in an expression vector, and the one chosen was pPICZαC. It contains, upstream of the multiple cloning site (MCS), a sequence coding for the *S. cerevisiae* α-factor secretion signal that, when maintained in the final protein product, leads to secretion of the recombinant protein. Since *P. pastoris* secretes only low levels of endogenous proteins, secretion of heterologously expressed proteins can be regarded as a first step of protein purification. A secreted protein can then be purified from the culture medium either by precipitation and centrifugation, or using an affinity column. Downstream of the MCS, there are sequences coding for the *c-myc* epitope and for a 6xHis tag. While the former may be useful for assessing protein expression levels by Western blot, the latter can also be used for affinity purification of the recombinant protein. The promoter driving the expression in pPICZαC is the tightly-regulated, methanol-induced *AOX1* promoter. ORF P1192R was PCR-amplified using L60 genomic DNA as template and cloned in pPICZαC.

Once the pPICZαC-P1192R clone was successfully obtained, it had to be inserted into *P. pastoris* cells. Since the pPICZαC vector does not contain a yeast origin of replication, transformants could only be isolated if recombination occurred between the plasmid and the cellular genome. To promote homologous recombination over random insertion of the plasmid, pPICZαC-P1192R was linearized using the restriction enzyme *Age*I prior to transformation. This enzyme hydrolyzes the plasmid just after the end of the ORF containing P1192R, the secretion factor and the fusion tags. X33 cells were then transformed by electroporation with the linearized plasmid and transformants were selected for zeocin resistance. They were also screened for the Mut phenotype, by picking and streaking them sequentially in MM and YPDS plates. This order was followed because streaking first in YPDS could leave traces of glucose in the tip that could inhibit the methanotrophic metabolism on the MM medium. In addition to screening for the Mut phenotype, streaking colonies to YPDS medium allows for confirmation of transformation, since some of the

colonies resulting from the transformation are just a result of transient expression of the zeocin-resistance gene. From the screening for the Mut phenotype, $Mut^+$ colonies (capable of utilizing methanol as a carbon source at normal levels, and therefore having both AOX genes) were selected.

### 4.3.4 Attempts to express recombinant protein in *Pichia pastoris*

Colonies confirmed to be $Mut^+$ were then used for initial testing of protein expression. This was performed using BMMY medium containing 0.5% methanol, for a total incubation period of 96 hours. Due to natural evaporation of methanol as well as its consumption by the cells, it was necessary to regularly add methanol to maintain induction, and this was done with 24-hour intervals, at which 100% methanol was added to the culture for a final concentration of 0.5%. During induction, samples from each culture were taken at several intermediate time-points (6, 12, 24, 36, 48, 60, 72, 84 and 96 hours of induction), for posterior analysis. At the endpoint, both the cell pellets and the supernatants were collected for analysis. Though the recombinant protein supposedly contained a secretion signal and thus should accumulate in the culture medium, it was possible that it wasn't efficiently secreted and accumulated inside of the cells. In addition, analysis of the cellular protein contents could hint for possible degradation signs of protein present in the supernatants. As the secretion factor is cleaved upon protein secretion, the expected size of the recombinant protein was around 140 kDa. Also, since *P. pastoris* tends to glycosylate secreted proteins, the real molecular weight of the protein could be even higher than expected. However, when analyzing the culture supernatants at each time-point by SDS-PAGE followed by Coomassie staining, only faint bands of about 75 kDa and 48 kDa were detected at later times. Furthermore, we did not observe any expression of pP1192R in Western blots using a commercial anti-6xHis antibody. Thinking that the absence of expression could be due to an incorrect construction of the expression strain, the pPICZαC-P1192R plasmid was linearized again, but this time using the restriction enzyme *Pme*I. Contrary to *Age*I, the *Pme*I enzyme hydrolyzes the pPICZαC-P1192R plasmid in the sequence of the *AOX1* promoter, which enhances homologous recombination with the *AOX1* promoter region of the cellular genome. Cells were transformed and transformants were selected with increasing concentrations of zeocin, with colonies being obtained up to 500 µg/ml (the usual zeocin concentration is 100 µg/ml). The rationale behind this was to promote the insertion of more than one copy of the plasmid into the genome, since resistance to zeocin is stoichiometric (Higgins *et al.*, 1998). Thus, transformants selected using 500 µg/ml of zeocin should, theoretically, have at least five copies of the plasmid inserted in their genome. The objective was to get a high expression

level of the recombinant protein by increasing the number of copies of the gene to be expressed in the genome of the cells. This time the presence of the expression cassette in the genome was confirmed by PCR. All the same, we still did not detect any expression of pP1192R in *P. pastoris* after quite a few expression attempts, assessed as described above.

### 4.3.5 Heterologous expression of P1192R in *Saccharomyces cerevisiae*

Because the attempts at expressing and/or purifying functional pP1192R from *E. coli* or *P. pastoris* cells failed, we turned to *S. cerevisiae* to accomplish our goal. One advantage of using this organism as an expression system was that many of the well-known and characterized eukaryotic type II DNA topoisomerases have been heterologously expressed in it (Worland & Wang, 1989; Austin *et al.*, 1995; Lavrukhin *et al.*, 2000; Dickey, Choi, Van Etten, & Osheroff, 2005) and, therefore, it had been shown to be useful for purposes similar to ours. It has also been, for more than two decades, a widely used expression system for the production of foreign proteins derived from a wide range of species. To express pP1192R in this system we used strain BY4741, a wild-type strain frequently used for research. ORF P1192R was cloned in the yeast expression plasmid pYES2 under the control of the yeast *GAL1* promoter. In a time-course analysis of total protein extracts of pYES2-P1192R transformants, upon induction of expression with galactose, we could detect, by Coomassie staining, the appearance and accumulation of a band slightly below the 135 kDa marker, approximating the predicted molecular mass for pP1192R (Figure 35). This suggested that, contrary to what we had observed for *P. pastoris*, expression of P1192R in *S. cerevisiae* was feasible and easily achieved.

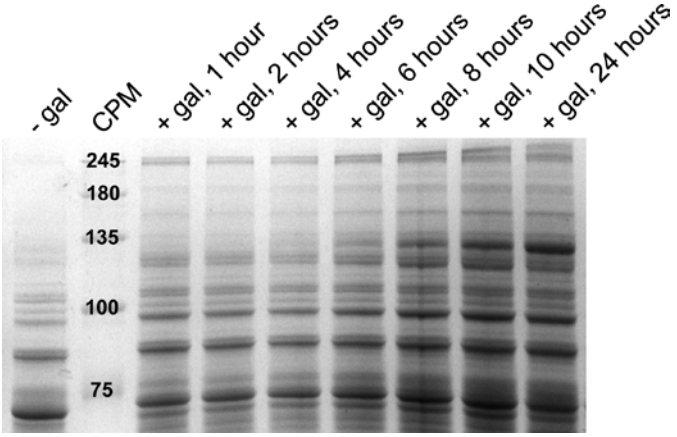


**Figure 35.** Time-course of pP1192R expression in *Saccharomyces cerevisiae*.
Total protein extracts, prepared as described in Section 4.2.10, from pellets of 50 ml cultures incubated at 25 ºC for the indicated periods (above each lane), were analyzed by SDS-PAGE followed by Coomassie staining. Protein expression was induced by addition of 2% galactose (+ gal). CPM is the molecular weight marker.

### 4.3.6 Expression and purification of pP1192R

Since the pYES2-P1192R clone did not contain a tag allowing for affinity purification, we cloned P1192R in pYES2 with the inclusion of a 6xHis tag. This was done using the pPICZαC-P1192R clone as template for PCR, because it already included the tag. Cells of strain BY4741 were then transformed with the pYES2-P1192R_6xHis clone. High expression levels of pP1192R_6xHis were observed after induction with galactose, in preliminary expression attempts using *S. cerevisiae* strain BY4741. Several conditions were then tested (incubation temperature, induction time, galactose concentration, number of cells at the time of induction), resulting in the choice of a 23 - 24 hour induction period at 25 $^{o}$C with 2% galactose. This was followed by solubility testing, in which the concentration and presence or absence of components of the lysis buffer was tested. We determined that the presence of glycerol [at 10% (v/v)] and of a high concentration of NaCl (1 M) were required to promote the solubility of the majority of expressed protein. With lower salt concentrations most of the protein would be retained in the insoluble fraction and in the absence of NaCl only traces of soluble pP1192R_6xHis were obtained. Since we were able to obtain soluble pP1192R under native conditions, inclusion of detergents in the lysis buffer was not tested at this point. Purification of the recombinant protein by affinity chromatography in a Ni NTA column using the 6xHis tag was initially attempted, successfully, using high concentrations (300 mM) of imidazole for its elution (Figure 36). However, no activity was observed for the eluted

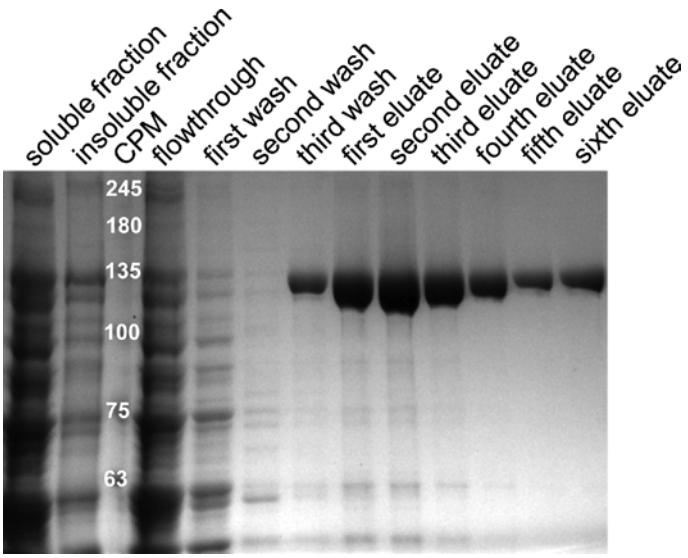


**Figure 36.** Preliminary purification of recombinant pP1192R_6xHis expressed in *S. cerevisiae*.
A soluble protein extract was prepared from 500 ml of culture of BY4741 cells expressing pP1192R_6xHis essentially as described on section 4.2.15. After binding of the protein to the Ni NTA resin, washes with 20, 50 and 100 mM imidazole were performed, and the protein was eluted using 300 mM imidazole. Samples of each fraction were analyzed by SDS-PAGE followed by Coomassie staining. CPM is the molecular weight marker.

protein in a decatenation (using kDNA) *in vitro* assay (Figure 37). We hypothesized that the absence of activity could be due to the presence of imidazole in the reaction, since imidazole-derived compounds are known to inhibit type II topoisomerases (Vashisht Gopal & Kondapi, 2001; Baviskar *et al.*, 2011), and this was tested using commercial human TopoIIα (hTopoIIα), which was indeed inhibited by 300 mM imidazole in our assays. We thus determined the lowest concentration of imidazole at which pP1192R_6xHis would elute from the column and found that 150 mM would suffice, and this concentration did not inhibit topoisomerase II activity in *in vitro* assays using the commercial hTopoIIα. Still, purified recombinant pP1192R_6xHis remained inactive. Dialysis of the purified protein was also attempted in order to completely remove the imidazole from the final protein solution, but it was always observed that during dialysis most of the protein precipitated out of solution, with only a small fraction remaining soluble. In addition, and contrary to what is suggested by several purification protocols (Biersack, Jensen, & Westergaard, 1999; Lindsley, 1999; Lavrukhin *et al.*, 2000; Dickey *et al.*, 2005) in which ours was based on, pP1192R_6xHis was found to precipitate at concentrations above 100 µg/ml. *In vitro* functional assays were attempted using either eluates in which the protein had not precipitated or soluble fractions of purified pP1192R_6xHis (directly or after a dialysis step). Several concentrations of protein were tested in the final reaction solution (up to 1 mM of pP1192R_6xHis) but decatenation activity was never detected.

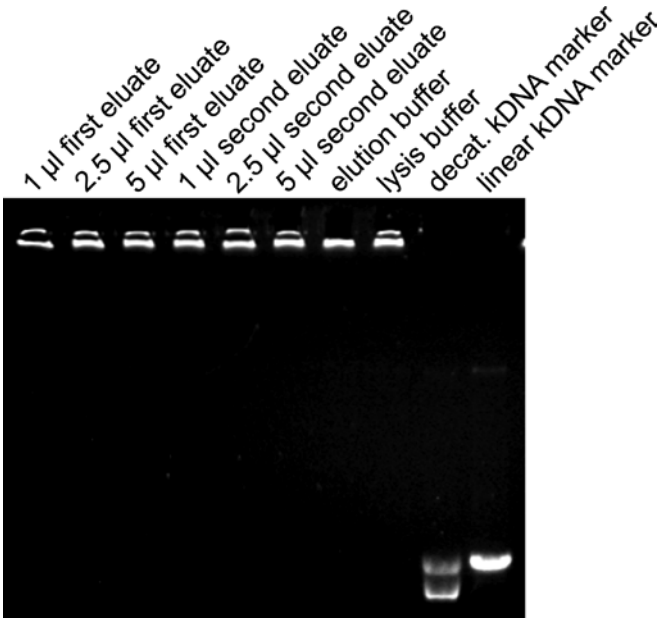


**Figure 37.** *In vitro* decatenation assay using purified pP1192R.
Several volumes of two eluates containing purified pP1192R were tested to assess for topoII-specific decatenation activity, using kDNA as substrate. Decatenated (decat.) and linear kDNA markers (TopoGEN) were used as indicators of topoII or nuclease activity, respectively.

Addition of a glutathione S-transferase (GST) tag for affinity purification of proteins has also been reported to promote recombinant protein solubilization (Esposito & Chatterjee, 2006; Costa, Almeida, Castro, & Domingues, 2014). Since we were having problems in maintaining purified pP1192R_6xHis soluble, we reasoned that expressing P1192R with a GST tag could prevent its precipitation after purification. For this, we substituted the 6xHis tag of clone pYES2-P1192R_6xHis with GST, and thus GST was fused to the C-terminus of P1192R. We then expressed pP1192R_GST and extracted it using the same conditions as for pP1192R_6xHis, but found that its solubility was not augmented in comparison to the latter. Additionally, it also did not aid in obtaining functional recombinant protein, as judged by *in vitro* activity assays.

### 4.3.7 Production of polyclonal anti-pP1192R serum using recombinant protein expressed in *S. cerevisiae*

Even though purified pP1192R was not active in our *in vitro* assays, we were able to produce significant amounts of protein and decided to use it for production of an anti-pP1192R polyclonal serum. Two CD1 mice were inoculated with the protein and their blood was collected at the end of the immunization protocol. The serum was tested for Western blot and immunofluorescence. In the first case, recombinant pP1192R expressed both in yeast and in *E. coli* was recognized as a band just below the 135 kDa marker, as expected (Figure 38), even though the bacterial extract was thoroughly recognized by the serum. Viral pP1192R was also recognized in total protein extracts of Ba71V-infected cells (as described on Chapter 3). The serum also worked in immunofluorescence labeling of Vero cells infected with Ba71V and of swine macrophages infected with L60 (as shown also on Chapter 3).

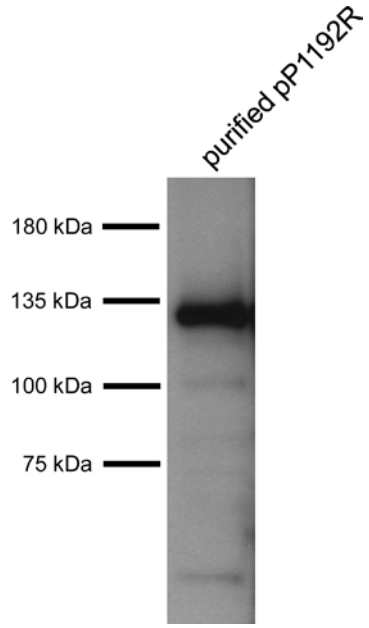


**Figure 38.** Western blot testing of the anti-pP1192R polyclonal serum.
Purified recombinant pP1192R, expressed in *S. cerevisiae* and used as inoculum for the production of the anti-serum, analyzed by Western blot using the produced polyclonal anti-pP1192R serum.

### 4.3.8 Expression and purification of Top2p

The failure in obtaining purified soluble pP1192R active in *in vitro* functional assays could be due to an error in the extraction and/or purification protocols. Up to now our control had been only for the *in vitro* assays *per se* and consisted on the use of a commercial purified human topoIIα protein, so we decided to develop a new control. Because we were expressing pP1192R in *S. cerevisiae*, and this organism codes for a single, well characterized type II topoisomerase, we considered it to be a good control for our method. We cloned the respective ORF, *TOP2*, in its wild-type version, in the expression vector pYES2 (with a sub-cloning step in pET24a to add a 6xHis tag). We then expressed, extracted and purified Top2p and pP1192R in identical conditions and found that the two enzymes required different extractions conditions. While pP1192R, as described before, was efficiently extracted using 1 M NaCl and 10% glycerol, solubilization of Top2p was inefficient under these conditions. Inversely, using the optimal conditions for Top2p, which involved the presence of 500 mM KCl and 10 mM $(NH_4)_2SO_4$ in the extraction buffer, pP1192R remained mostly insoluble. Thus, we purified both enzymes using the optimal conditions for each and we performed *in vitro* activity assays to check for activity. We observed topoisomerase II activity for Top2p when it was purified under optimal conditions (Figure 39) but, as before, no activity was

observed for pP1192R. We could also observe that excess Top2p enzyme in the reaction could mimic absence of activity, as described in the literature (Nitiss *et al.*, 2012).

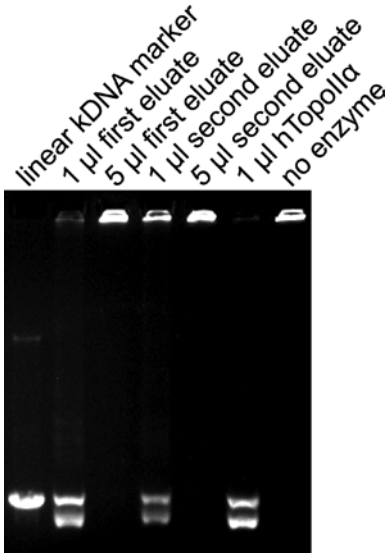


**Figure 39.** *In vitro* decatenation assay using purified Top2p.
Several volumes of two eluates containing purified Top2p were tested to assess for type II topoisomerase-specific decatenation activity, using commercial kDNA as substrate. Commercial human TopoIIα (hTopoIIα) was used as a positive control. Linear kDNA marker (TopoGEN) was used as indicator of possible nuclease activity.

### 4.3.9 Functional complementation of Top2p thermo-sensitive *S. cerevisiae* mutants

Considering the above described unsuccessful attempts to demonstrate pP1192R's topoisomerase activity *in vitro* and, in addition, the existence of *S. cerevisiae* thermo-sensitive mutants of the endogenous topoisomerase II, coded by the *TOP2* gene, we decided to perform complementation assays for an indirect proof of function. We could use strains JCW26 and SD117 (Wasserman *et al.*, 1993), which contain a thermo-sensitive mutation in *TOP2* (*top2 ts*) that sustains growth at 25 ºC (permissive temperature), but makes Top2p non-functional at 35 ºC (restrictive temperature). We transformed cells of these thermo-sensitive strains with plasmid pYES2-P1192R and checked for functional complementation at the restrictive temperature, in the presence of galactose, without success. From our results in mammalian culture cells, we had the indication that pP1192R is a strictly cytoplasmic protein. Seeing as in *S. cerevisiae* cells Top2p is a nuclear protein and its substrate (the DNA) is also located in the nucleus, and because in *S. cerevisiae* the nuclear envelope remains intact during cell division (Arnone, Walters, & Cohen-Fix, 2013), we hypothesized that the absence of functional complementation could simply be due to the impossibility of pP1192R to act upon the yeast DNA. Top2p has been extensively studied and

its nuclear localization signals have been well characterized and shown to be present in the C-terminal domain (Caron, Watt, & Wang, 1994; Jensen *et al.*, 1996). Therefore, we fused pP1192R with the Top2p C-terminal domain (from amino acid position 1189 to position 1428 of Top2p; see Figure 40), expecting that we could force pP1192R to enter the nucleus of *S. cerevisiae* cells. To verify this, we also constructed fusions of each recombinant clone with GFP, so that we could track the localization of the protein inside of the cell after its expression. Our expectation was confirmed by fluorescence microscopy, since in cells expressing GFP-pP1192R fused with the Top2p C-terminal domain, the fusion protein is

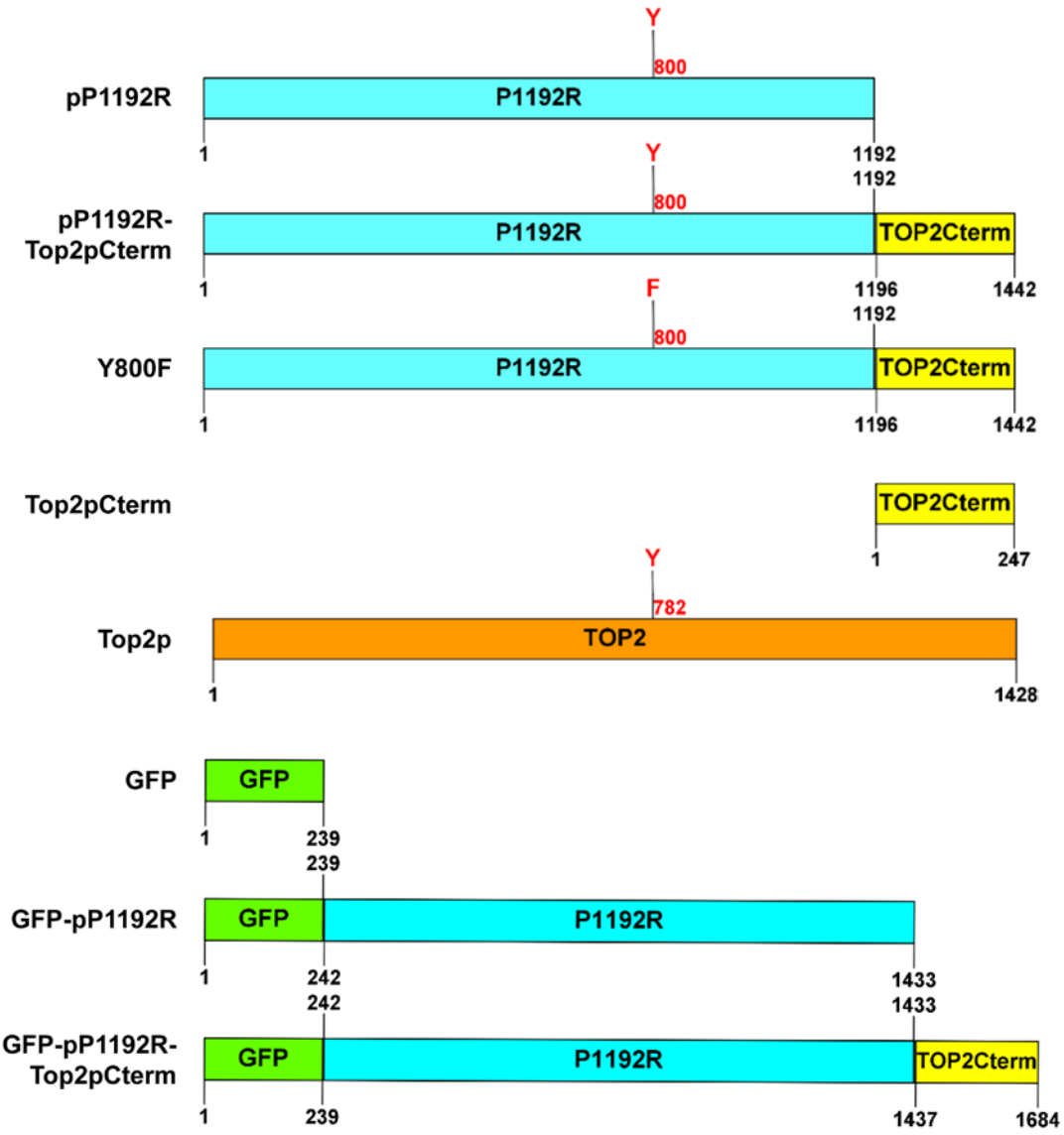


**Figure 40.** Schematic representation of the recombinant proteins expressed in complementation assays.

Constructs in which GFP is absent were used for complementation assays. In these, the catalytic tyrosine is indicated at position 800 for pP1192R or position 782 for Top2p. In the pP1192R catalytic mutant, the tyrosine was substituted by a phenylalanine. GFP-containing constructs were used for localization assays. The starting and ending residues of each domain/protein in the final construct are indicated.

indeed present exclusively in the nucleus, while in cells expressing GFP-pP1192R this protein is localized solely in the cytoplasm and GFP alone is seen having both nuclear and cytoplasmic localization (Figure 41).

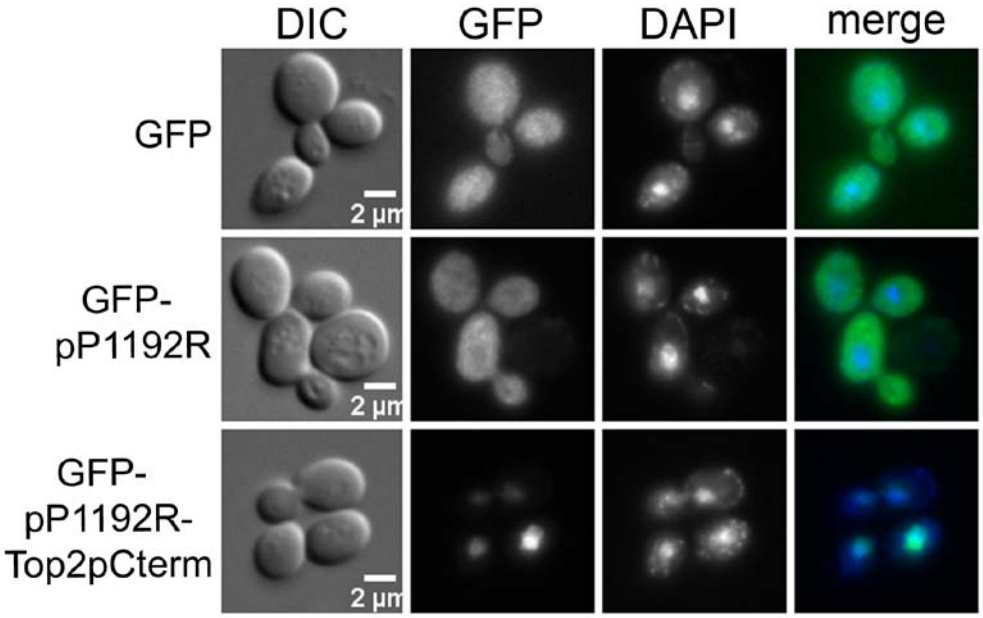


**Figure 41.** Localization of GFP-fusion constructs in *S. cerevisiae* cells.
Yeast strain JCW26 was transformed with pYES2-GFP, pYES2-GFP-P1192R or pYES2-P1192R-TOP2Cterm. Cultures of each transformant were diluted to a starting $OD_{600}$ of 0.5 in galactose-containing medium and incubated at 25 ºC for 16 hours. In the merged panels, the GFP signal is shown in green and DNA DAPI staining in blue. Experiments were performed at least twice and representative images are shown.

We tested again for functional complementation and found that, in opposition to pP1192R, the pP1192R-TOP2Cterm fusion protein rescues a *top2 ts* mutation (Figure 42). This rescue was even more evident in liquid cultures, in which we observed that expression of the pP1192R-TOP2Cterm chimera in cells grown at the restrictive temperature would allow for growth, while cells containing only the empty expression vector, expressing the wild-type pP1192R or expressing just the Top2p C-terminal domain didn't grow at the restrictive temperature. Additionally, in the pP1192R-TOP2Cterm fusion we mutated the predicted catalytic residue, a tyrosine at position 800, to a phenylalanine (Figure 40), in order to abolish its activity and, thus, the rescue of the JCW26 strain at the restrictive temperature. As expected, no rescue was observed for the catalytic mutant (Figure 42), both in solid and in liquid media. This confirms the importance of the predicted catalytic residue for activity and substantiates the function of pP1192R as a type II topoisomerase. Similar results were obtained when we used another *top2 ts* strain, SD117 (Wasserman *et al.*, 1993).

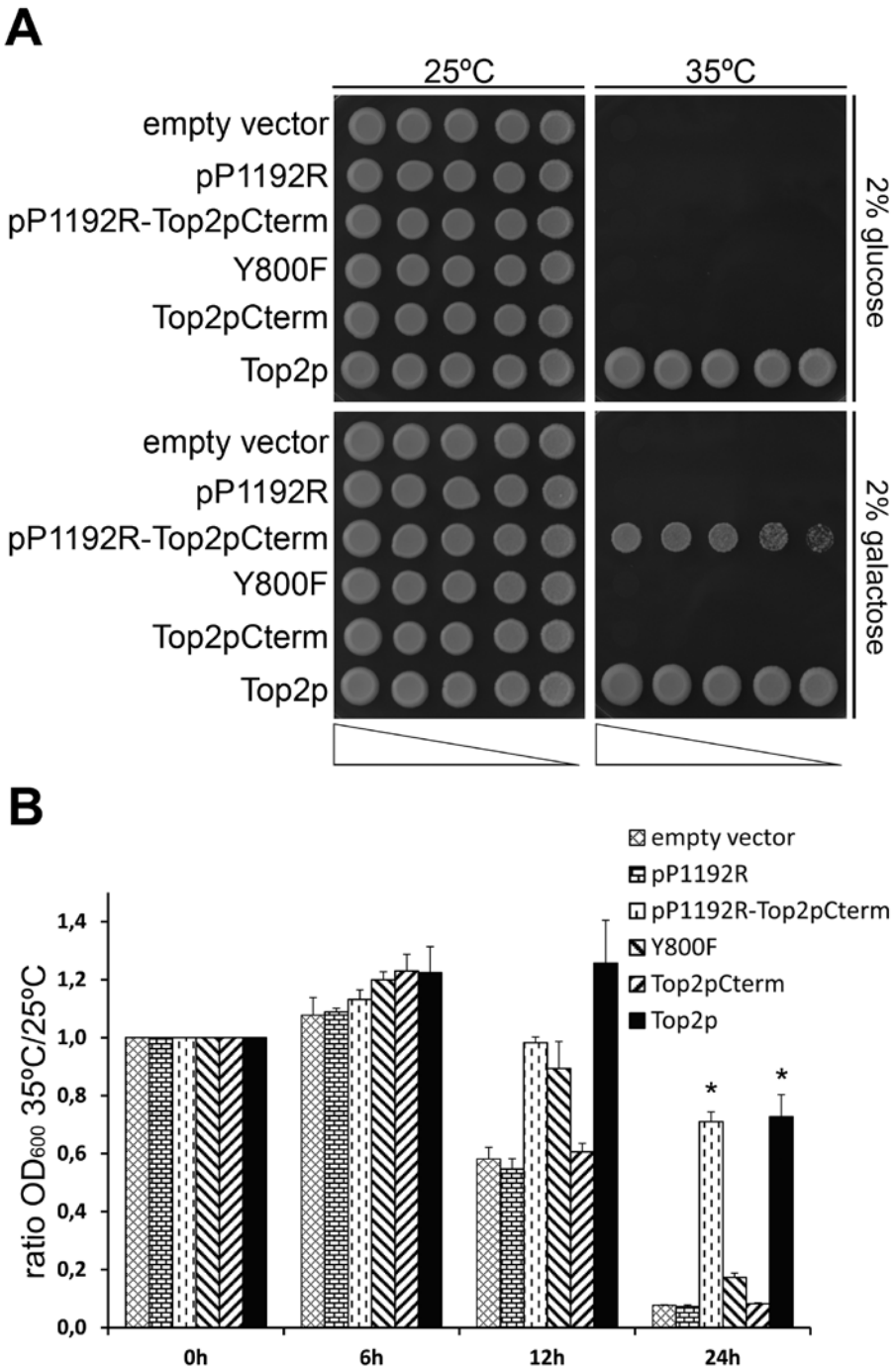


**Figure 42.** Functional complementation of a yeast *top2* temperature-sensitive mutation by P1192R. (A) Yeast strain JCW26 was transformed with pYES2, pYES2-P1192R, pYES2-P1192R-TOP2Cterm, pYES2-P1192R-TOP2Cterm Y800F, pYES2-TOP2Cterm or pYES2-TOP2. Serial three-fold dilutions of each transformant were plated on YP media containing glucose or galactose and incubated at either 25 ºC or 35 ºC. The plates were photographed on the fifth day of incubation. The image is representative of, at least, three independent experiments. (B) Yeast cultures of transformants as in (A) were diluted to a starting $OD_{600}$ of 0.1 in galactose-containing medium and then split in half and incubated at either 25 ºC or 35 ºC. The graphic represents the mean (± standard error) of the ratio between the $OD_{600}$ values obtained at the indicated time-points for the cultures incubated at 35 ºC and the cultures incubated at 25 ºC, for four different colonies for each transformant, from three independent experiments. The asterisk marks results that are significantly different ($p < 0.05$) using a one-way analysis of variance (ANOVA) and post hoc tests.

### 4.3.10 *In vitro* demonstration of pP1192R topoisomerase II activity

It could still be argued that pP1192R was only active because it was fused to the C-terminal domain of the yeast Top2p, and that the activity of pP1192R *per se* remained to be demonstrated. As topoII *in vitro* activity assays were performed at 37 ºC, temperature at which the thermo-sensitive Top2p from yeast strain JCW26 should not be functional, we assessed for pP1192R activity in total yeast extracts of JCW26 transformants. We used a decatenation assay, since type II topoisomerases are unique in their ability to catalyze the decatenation of double-stranded DNA molecules, with kDNA from *Crithidia fasciculata* and conditions adapted from Adachi *et al.* (1992), in which a similar assay was performed to demonstrate activity for the mouse topoisomerase II. In initial experiments we observed activity in extracts of Top2p-expressing cells but not in pP1192R-expressing protein extracts. This presence/absence of activity correlated well with the solubility of the respective topoisomerase, with Top2p being efficiently solubilized under the conditions used while the majority of pP1192R was retained in the insoluble fraction. Addition of Triton X-100 to the extraction buffer promoted the solubility of recombinant pP1192R, or perhaps avoided its precipitation, and using these optimized conditions for extraction, we tested increasing amounts of total extracts of JCW26 cells expressing pP1192R, a pP1192R catalytic mutant, the pP1192R-Top2pCterm fusion or the wild-type Top2p, as confirmed by SDS-PAGE (Figure 43), and observed that pP1192R was able to decatenate kDNA as efficiently as

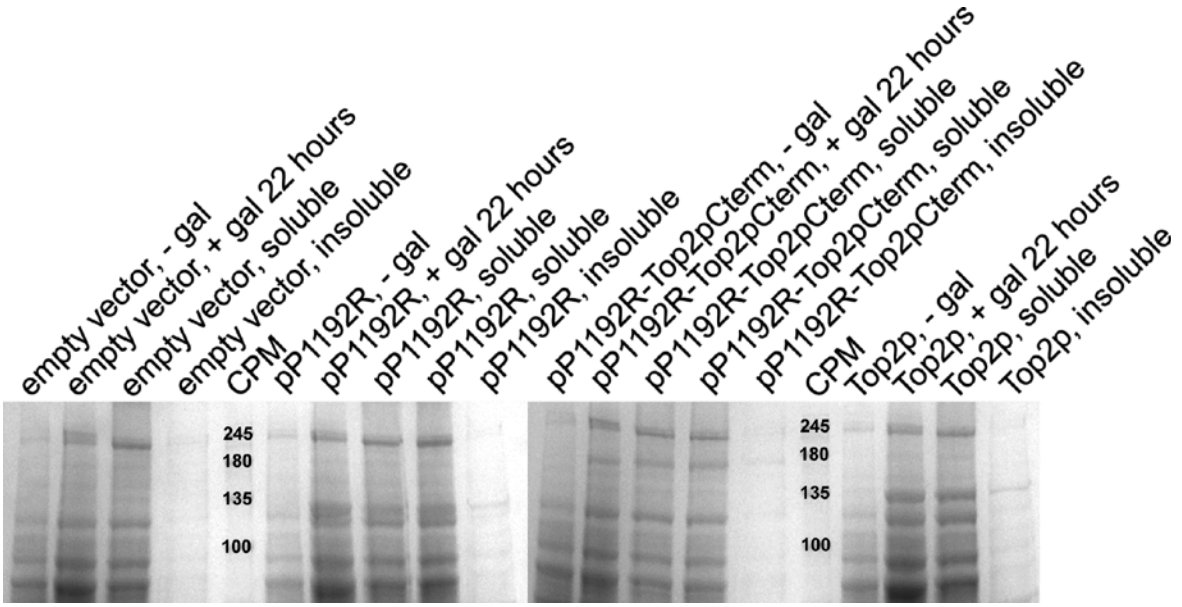


**Figure 43.** SDS-PAGE analysis of total protein yeast extracts.
Yeast strain JCW26 was transformed with pYES2, pYES2-P1192R, pYES2-P1192R-TOP2Cterm or pYES2-TOP2. After induced expression of the respective recombinant proteins with 2% galactose over a 22-hour period, total yeast extracts were prepared and analyzed by SDS-PAGE followed by Coomassie staining. CPM is the molecular weight marker.

wild-type Top2p (Figure 44). Validating previous results, mutation of the predicted catalytic residue fully abrogated the activity of pP1192R. The activities detected were all ATP-dependent, as expected for type II topoisomerases.

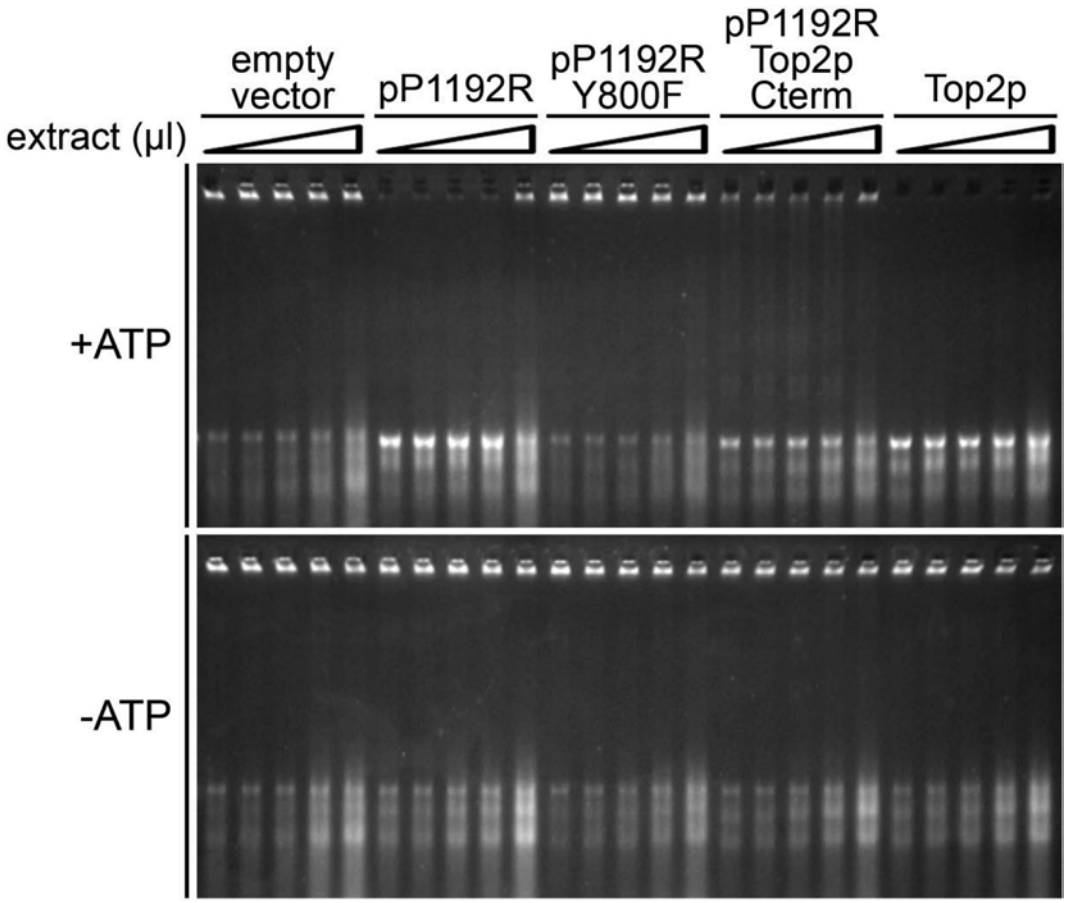


**Figure 44.** Decatenation activity *in vitro* assay using total protein yeast extracts.
Increasing amounts (0.1, 0.25, 0.5, 1.0 and 2.0 µl) of each soluble extract shown in A), as well as similar extracts of JCW26 transformed with pYES2-P1192R Y800F, were tested in an assay using kinetoplast DNA as substrate. For both panels, representative images of two independent experiments are shown.

### 4.3.11 Purification of functional recombinant pP1192R

Having successfully obtained active recombinant pP1192R, we decided to attempt a new purification after expression in yeast cells. To confirm that the difference between obtaining purified soluble protein and purified precipitated protein was in the addition of Triton X-100, we used the extraction conditions that were initially found to allow purification of high amounts of recombinant protein (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10% glycerol, 1 mM PMSF) with and without the addition of Triton X-100 for a final concentration of 0.1%. We found that Triton X-100 at this concentration was indeed essential to obtain recombinant

protein that did not precipitate after purification and it allowed for activity of the purified protein. We further investigated if Triton had to be present since the extraction buffer or if it could be added only during the washes, after binding of the protein to the Ni NTA resin. Addition of Triton only during the washing steps still allowed for purification of active pP1192R, even though the number of eluate fractions in which topoII activity was detected was lower. Thus, the best option was to add Triton X-100 at 0.1% to the extraction buffer and maintain the detergent present throughout the purification procedure.

We then used these conditions in a new purification of pP1192R. Following purification, eluate fractions were assessed for relaxation activity and the peak of activity coincided with the elution peak of recombinant pP1192R (Figure 45). Eluates found to contain high levels of topoII activity were pooled, quantified, aliquoted and stored at -80ºC until being used.

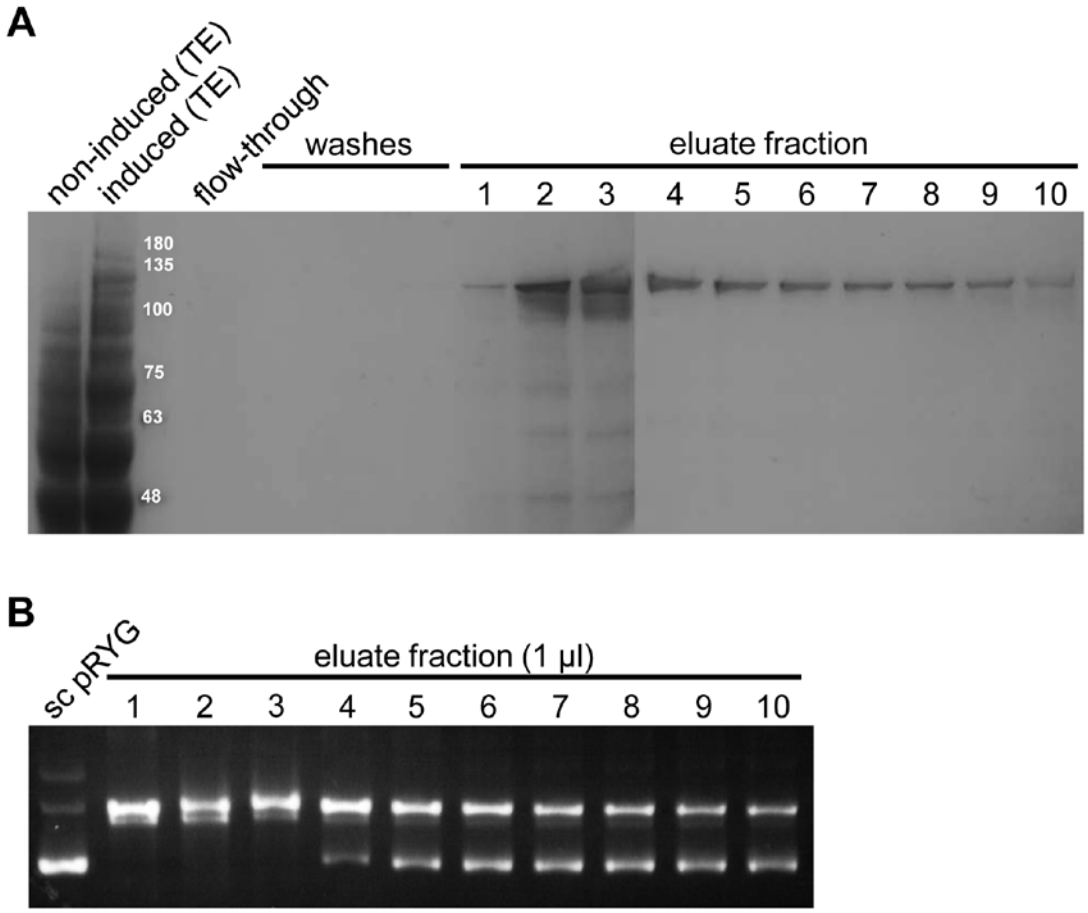


**Figure 45.** Purification of functional recombinant pP1192R.
A) Soluble functional protein was purified from 250 ml of culture of BY4741 cells expressing pP1192R_6xHis as described on section 4.2.20. After binding of the protein to the Ni NTA resin, washes with 20, 50 and 75 mM imidazole were performed, and the protein was eluted using 250 mM imidazole. 20 µl of each fraction (indicated above the lanes) were analyzed by SDS-PAGE followed by Coomassie staining. TE is an abbreviation for total extract.
B) Equal amounts (1 µl) of each eluate (indicated above each lane) shown in A) were tested for topoisomerase activity in a plasmid relaxation assay using supercoiled (sc) pRYG as substrate.

According to the quantification, performed using the Bradford protein assay in triplicate and assuming 10% of contaminants in the final pool, the protein was purified at a concentration of 4.43 µM, which means that the expression/purification procedure produced 1.044 milligrams of active pP1192R from about 250 ml of culture. The amount of protein to be used for characterization of its activity in *in vitro* assays was determined by performing relaxation and decatenation assays with dilutions ranging from 1 nM to 100 nM (Figure 46). We sought to use sub-optimal amounts since these would allow observing improvement effects upon variation of the conditions during testing. Therefore, for relaxation assays 5 nM of pP1192R were found to suffice, as were 10 nM for decatenation assays. As seen previously (Figure 44), pP1192R is able to recatenate decatenated kDNA when present in high concentrations.

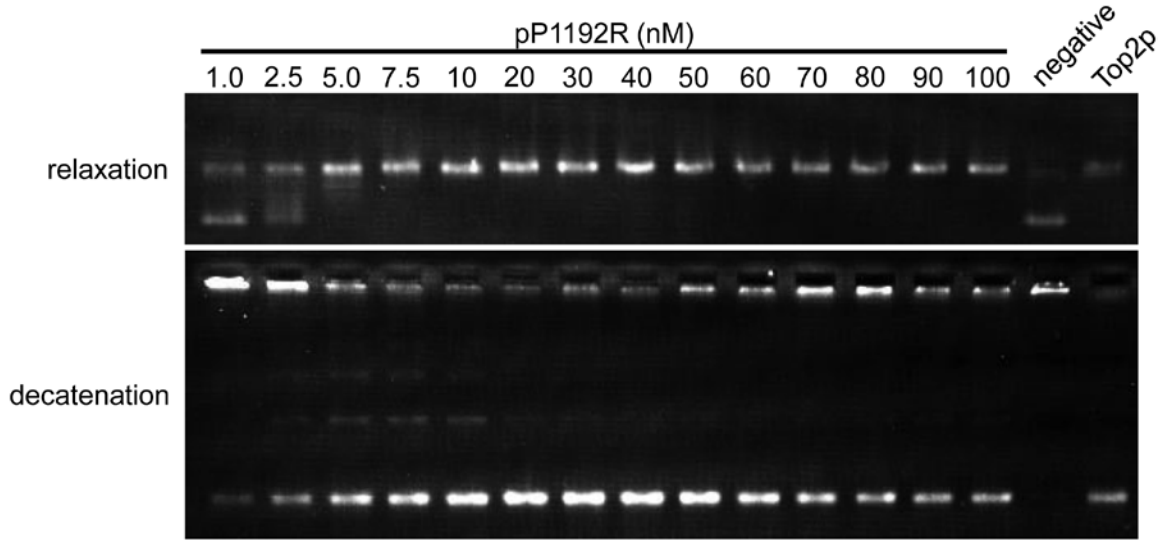


**Figure 46.** Determination of pP1192R's concentration to be used in *in vitro* assays for the characterization of its activity.
pP1192R was tested in increasing concentrations, ranging from 1 to 100 nM (final concentration indicated above each lane), in parallel relaxation (top panel) and decatenation (bottom panel) *in vitro* assays, using supercoiled pRYG or kDNA as substrates, respectively. Reactions were incubated at 37 ºC for 30 minutes, after which they were stopped by addition of 4 µl of stop solution. Purified Top2p was used as a positive control, while for the negative control an identical volume of enzyme dilution buffer was added to the reaction instead of pP1192R. For both panels, images are representative of two independent experiments.

### 4.3.12 Characterization of pP1192R activity *in vitro*

Having confirmed that purified recombinant pP1192R possesses decatenation and relaxation activities, we now wanted to establish the conditions for which these activities are optimal. For this we chose to use the *in vitro* relaxation assay, as it would be easier to observe subtle differences in activity due to changes in the reaction conditions than if we used the decatenation assay. We varied a single condition at a time, while maintaining all the others

constant, to determine optimal salt, divalent cations and ATP concentrations, pH, and temperature for pP1192R topoisomerase II activity. For salt, use of NaCl or KCl was found to be optimal at a concentration of 75 mM (Figure 47), with slight deviations from the optimal concentration being better tolerated with KCl.

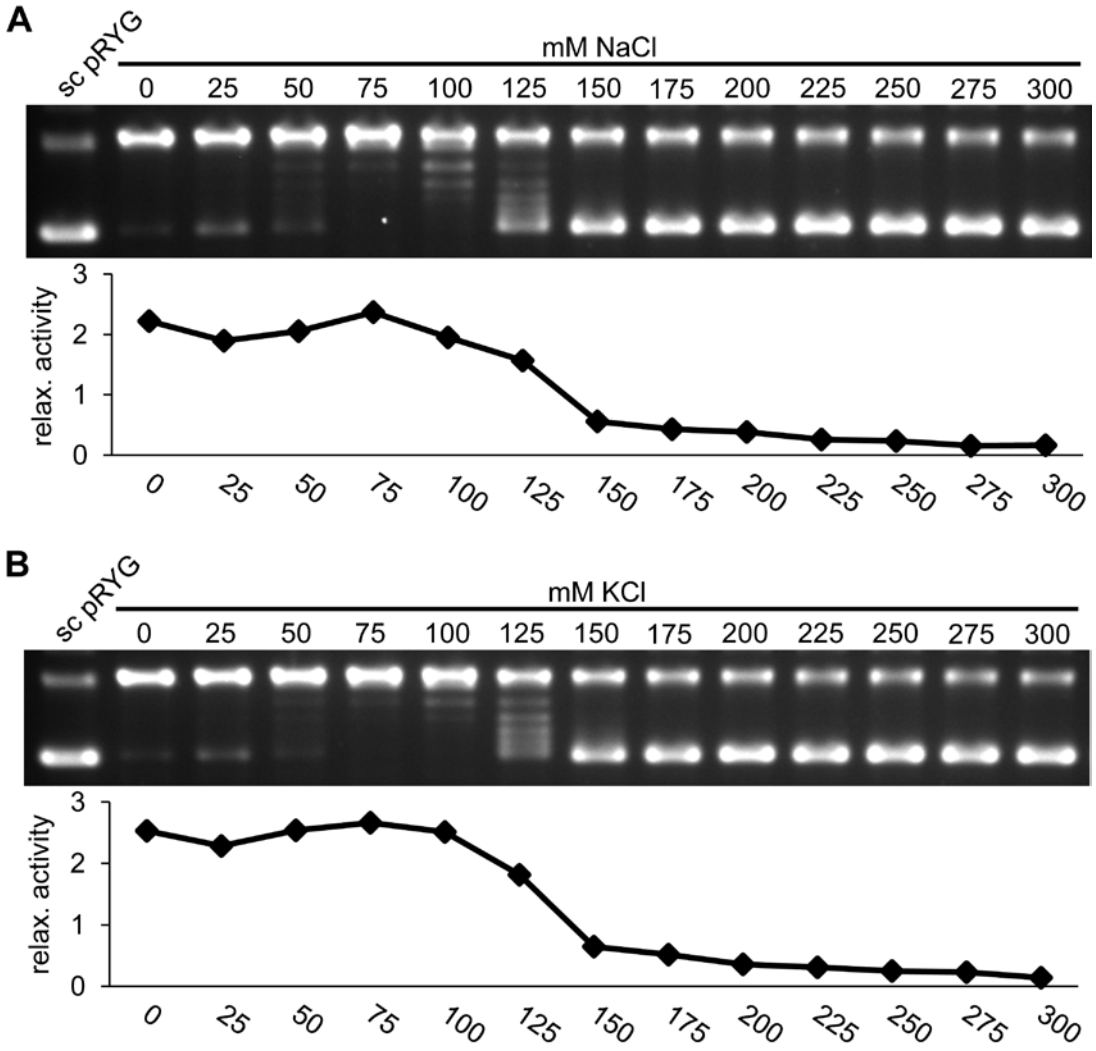


**Figure 47.** The effect of NaCl (A) or KCl (B) on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, using supercoiled pRYG as substrate. The final concentration used in each reaction is indicated at the top of each lane. Relaxation (relax.) activity of pP1192R was determined by densitometry quantification of the transition from supercoiled to relaxed forms and is expressed in relation to the control (first lane). For both panels, the results are representative of, at least, three independent experiments.

pP1192R's activity was highest between pH 7.0 and 8.0, with 7.5 being optimal, but the enzyme was almost inactive at pH 9.0 (Figure 48). Of note, the pH of the buffers used was set at 25 ºC and no attempt was made to correct for its variation due to temperature changes during the assays.

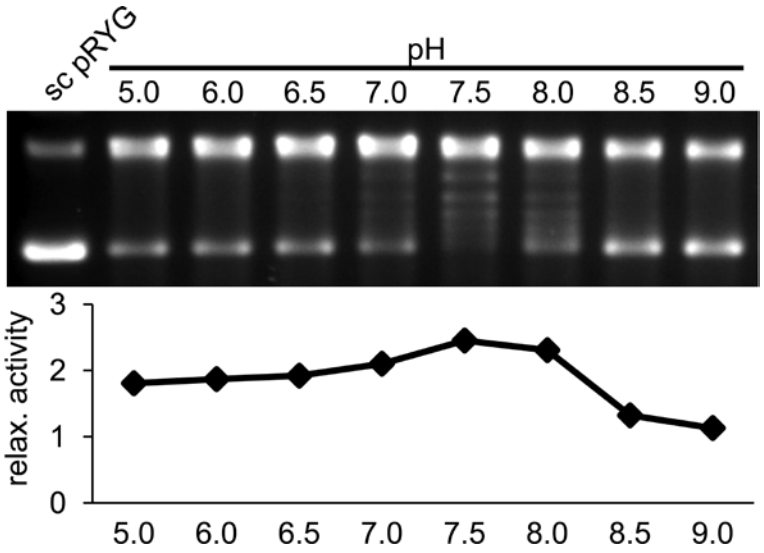


**Figure 48.** The effect of pH on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, except the pH was varied by adding to each reaction Tris-HCl buffer set (at 25 ºC) for the indicated pH (above each lane). Negatively supercoiled pRYG was used as substrate. Relaxation activity of pP1192R was determined by densitometry quantification of the transition from supercoiled to relaxed forms and is expressed in relation to the control (first lane). The result is representative of two independent experiments.

The optimal temperature for activity ranged between 30 and 37 ºC, and while 50 ºC still sustained some activity, no activity was observed below 20 ºC, for the reaction time tested (Figure 49).

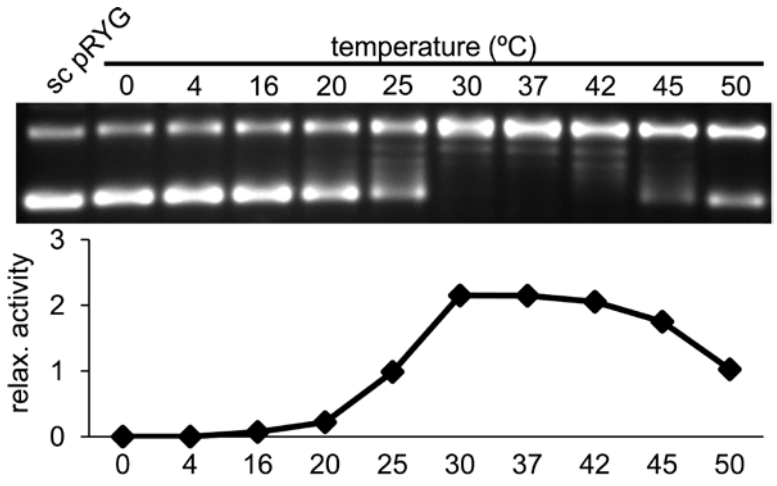


**Figure 49.** The effect of temperature on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, using negatively supercoiled pRYG as substrate. The temperature at which each reaction was incubated is indicated above each lane. Relaxation activity of pP1192R was determined by densitometry quantification of the transition from supercoiled to relaxed forms and is expressed in relation to the control (first lane). The result is representative of, at least, two independent experiments.

An increase in ATP concentration resulted in an increase of pP1192R activity, with at least 2 mM being required for optimal activity (Figure 50). Similarly to what has been observed for other topoisomerases (Osheroff, Shelton, & Brutlag, 1983; Huang, Wei, & Casjens, 1985; Peng & Marians, 1993; Lavrukhin *et al.*, 2000; Makarevitch & Somers, 2005), dATP efficiently substituted ATP in the reaction, since it can still be hydrolyzed to dADP + $P_i$, while dGTP did not support topoisomerase II activity.

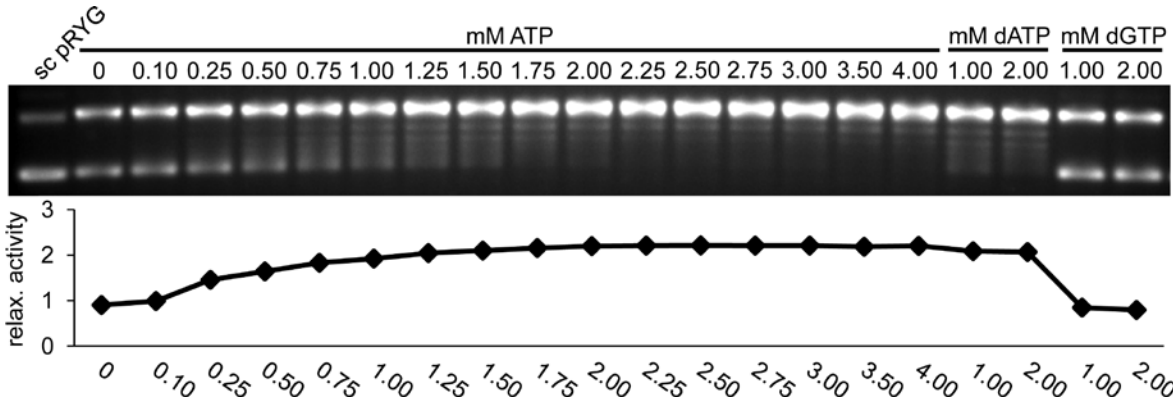


**Figure 50.** ATP dependence and the effect of ATP concentration on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, using negatively supercoiled pRYG as substrate. The final concentration used in each reaction is indicated at the top of each lane. Relaxation activity of pP1192R was determined by densitometry quantification of the transition from supercoiled to relaxed forms and is expressed in relation to the control (first lane). The result is representative of three independent experiments.

Finally, pP1192R required $Mg^{2+}$ at 6 mM for optimal activity, and at concentrations higher than 10 mM the activity was reduced (Figure 51A). Magnesium could be substituted by $Mn^{2+}$ (at least at 4 mM) or $Cu^{2+}$ (at 2 mM) but not by $Ca^{2+}$ or $Zn^{2+}$ (Figure 51B), and curiously $Mn^{2+}$ supported high relaxation levels even at 20 mM, without decrease in activity, contrary to what was observed for $Mg^{2+}$ or $Cu^{2+}$.

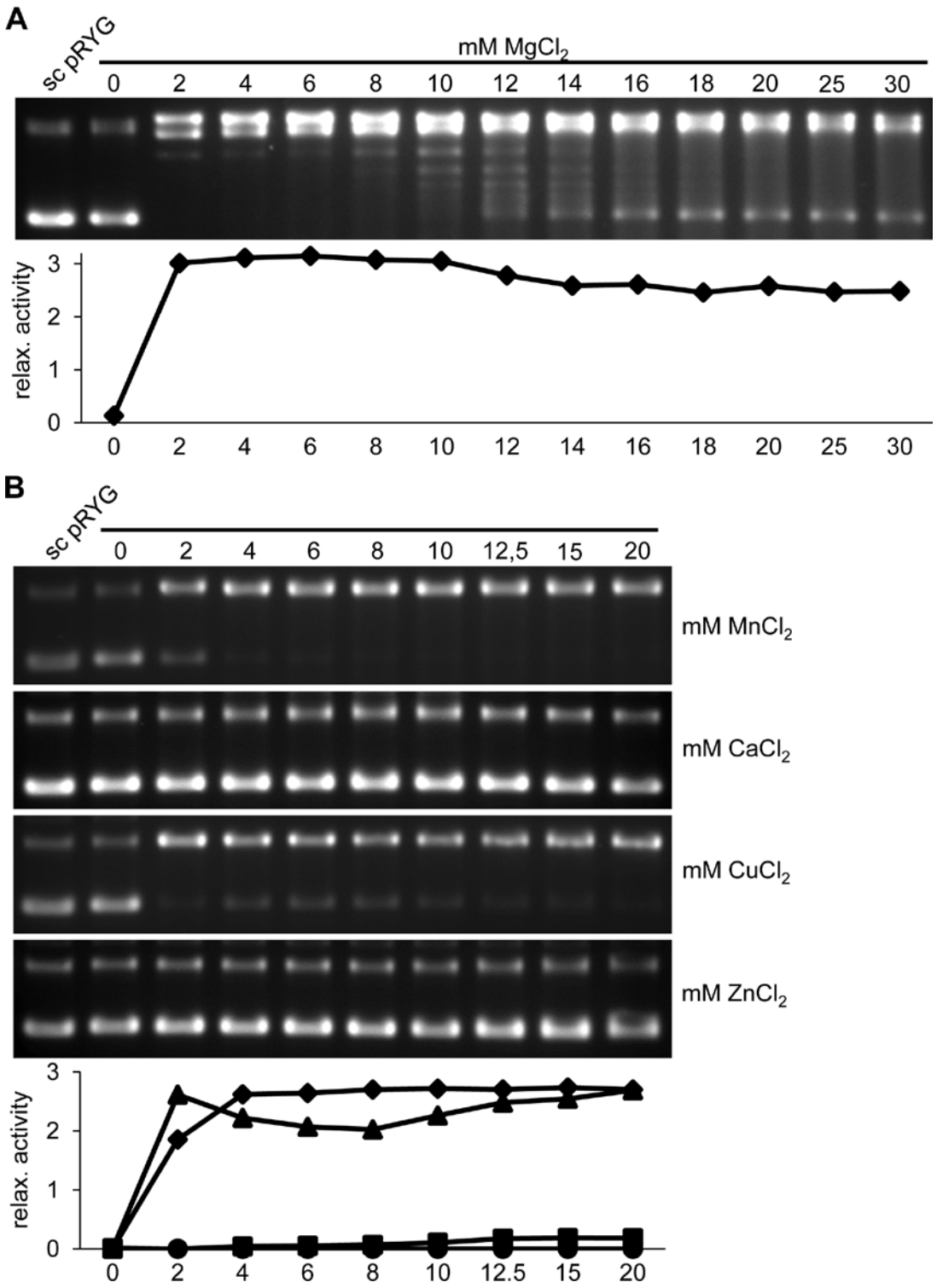


**Figure 51.** The effect of divalent ions on the relaxation of supercoiled plasmid DNA catalyzed by pP1192R.

*In vitro* relaxation reactions were performed as described in section 4.2.21, using negatively supercoiled pRYG as substrate. (A) Magnesium or (B) manganese (diamond), calcium (square), copper (triangle) and zinc (circle) were tested for their ability to support pP1192R's topoisomerase activity. The final concentration used in each reaction is indicated at the top of each lane. Relaxation activity of pP1192R was determined by densitometry quantification of the transition from supercoiled to relaxed forms and is expressed in relation to the control (first lane). Results are representative of, at least, two independent experiments.

Once we established the optimal conditions for pP1192R activity, these were used to assess if pP1192R worked in a distributive or in a processive fashion. A topoisomerase working in a processive mode binds to DNA and removes many, if not all, supercoils before dissociating from the substrate. On the contrary, when working in a distributive mode the enzyme removes only one or a few supercoils per binding event since the tendency for dissociation is higher (Osheroff *et al.*, 1983). As seen on Figure 52, initial substrate (supercoiled plasmid) and final reaction product (relaxed plasmid) are more abundant than intermediate forms, confirming that pP1192R is processive. Curiously, salt concentrations slightly higher than the optimal seem to switch pP1192R into a distributive mode (see Figure 47, 100 and 125 mM lanes), and this has also been observed for other type II topoisomerases (Osheroff *et al.*, 1983; Caron *et al.*, 1994; Lavrukhin *et al.*, 2000; McClendon, Rodriguez, & Osheroff, 2005).

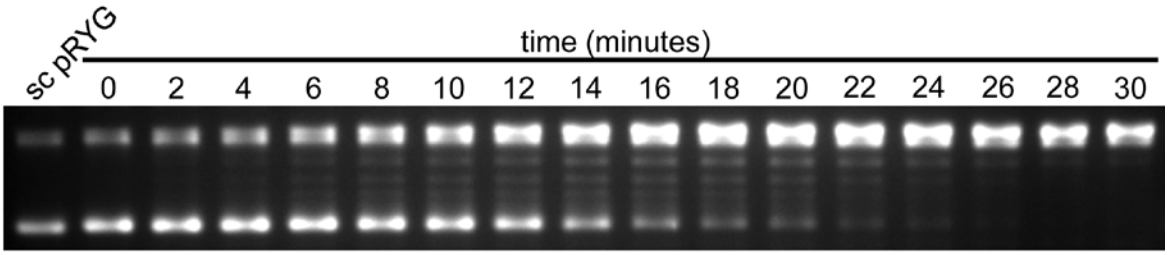


**Figure 52.** Time course of the relaxation of supercoiled DNA by pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, using negatively supercoiled pRYG as substrate. The time of incubation of each reaction is indicated above each lane. Reactions were stopped in two-minute intervals. Image is representative of two independent experiments.

We also investigated the ability of pP1192R to introduce supercoils in relaxed plasmid DNA, by using commercial relaxed pBR322. Either at working concentrations or at a clear excess in the reaction, pP1192R did not supercoil the relaxed plasmid (Figure 53). This result was anticipated as the bacterial DNA gyrase is the only topoII known to possess the ability of introducing supercoils in DNA molecules. Similarly to what we had previously observed (see Figure 46), when present at high concentrations pP1192R was able to catenate the relaxed plasmid molecules, which were retained in the well during agarose gel electrophoresis.

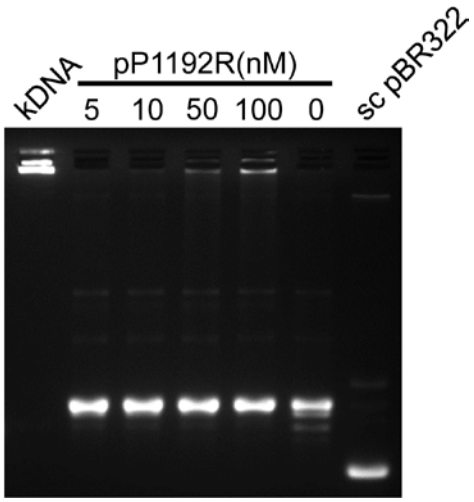


**Figure 53.** Testing the supercoiling ability of pP1192R.
*In vitro* relaxation reactions were performed as described in section 4.2.21, using relaxed pBR322 as substrate. The final enzyme concentration used in each reaction is indicated at the top of each lane. Supercoiled pBR322 was used as indicator of what a positive reaction should look like, while kDNA was used as a control for catenation. Image is representative of, at least, two independent experiments.

### 4.3.13 Inhibition of pP1192R by using anti-topoII compounds

Having characterized the *in vitro* activity of pP1192R, our next objective was to test several compounds known to impair DNA topoisomerases and assess if they had any effect on the activity of pP1192R. The compounds used in this work can be grouped into two categories: topoisomerase poisons – etoposide, *m*-AMSA, doxorubicin, genistein, nalidixic acid, ciprofloxacin, norfloxacin, enrofloxacin and gatifloxacin – which trap DNA-topoisomerase cleaved complexes and induce the accumulation of double-strand breaks in the DNA; and topoisomerase inhibitors – coumermycin A1, naringenin and topotecan - that do not induce the formation of cleaved complexes, but instead inhibit the topoisomerase from acting on the DNA. Curiously, doxorubicin is described as acting both as a poison, at low concentrations, and as an inhibitor, at high concentrations (Pommier *et al.*, 2010). Topotecan is a known inhibitor of type I topoisomerases. The topoII poisons used include antibacterial quinolones: nalidixic acid is a first generation quinolone, ciprofloxacin and norfloxacin are both second generation, enrofloxacin is of the third generation and gatifloxacin is a fourth generation quinolone. One major change from the first to the second generation of quinolones was the addition of a fluorine atom (Anderson & Osheroff, 2001) to the compounds, which conferred increased potency in comparison to first generation quinolones and lead to the alternative designation of fluoroquinolones.

Compounds were tested using decatenation assays, since we wanted to know their effect on the specific activity of a type II DNA topoisomerase. Of the non-quinolone agents tested

(Figures 54 and 55), coumermycin A1 and doxorubicin were found to inhibit pP1192R at very low doses, with 1 µM having already some effect. *m*-AMSA and genistein were effective at 8 and 32 µM, respectively, while higher doses were required for naringenin and etoposide (128 and 256 µM, respectively) to have some effect.

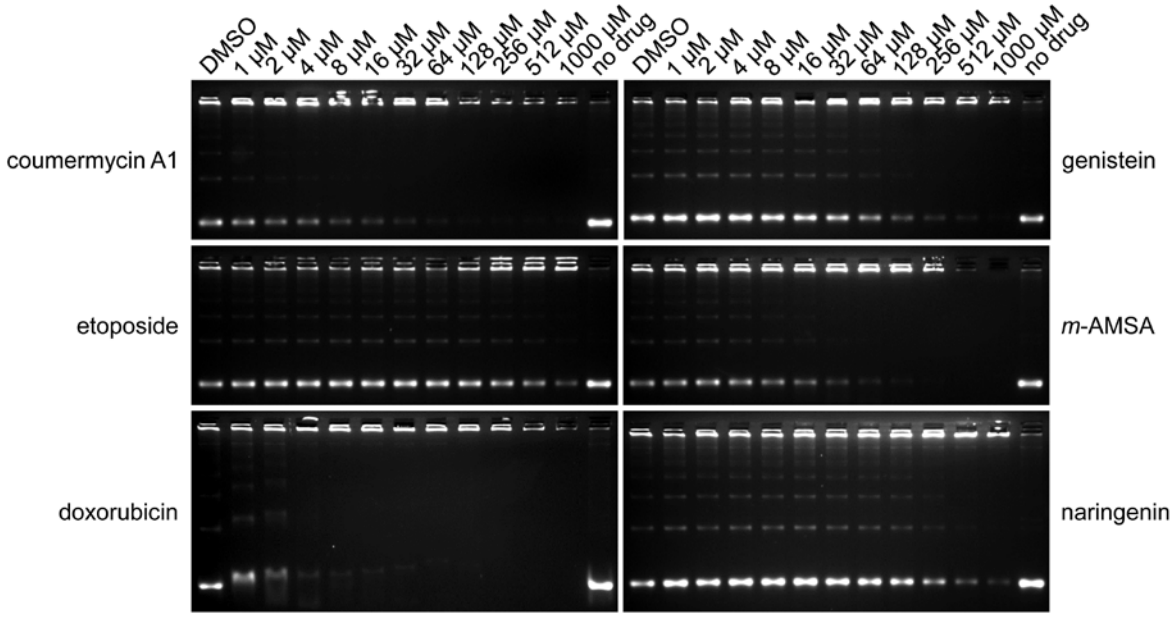


**Figure 54.** Inhibition of decatenation activity of pP1192R.
*In vitro* decatenation reactions were performed essentially as described in section 4.2.21, using kDNA as substrate and optimal conditions defined for pP1192R activity. Reactions were incubated at 37 ºC for 30 minutes, after which they were stopped by addition of 4 µl of stop solution. The final drug concentration used in each reaction is indicated at the top of each lane. Images are representative of, at least, two independent experiments.

Topotecan had no effect on pP1192R activity, up to 16 µM. However, at higher concentrations a highly fluorescent stain is present and obstructs visualization of the result.

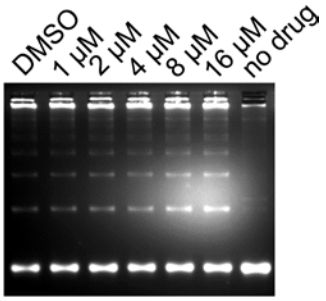


**Figure 55.** Testing the effect of topotecan on pP1192R activity.
*In vitro* decatenation reactions were performed essentially as described in section 4.2.21, using kDNA as substrate and optimal conditions defined for pP1192R activity. Reactions were incubated at 37 ºC for 30 minutes, after which they were stopped by addition of 4 µl of stop solution. The final drug concentration used in each reaction is indicated at the top of each lane. Image is representative of two independent experiments.

As for the quinolones tested, ciprofloxacin was the one for which a lower dose (128 µM) was required to observe an effect, and it was also the one for which a higher effect was observed at the highest concentration tested. Enrofloxacin was similar, but slightly less effective at higher doses, while norfloxacin and gatifloxacin were less effective (minimum dose being 256 and 512 µM, respectively). Nalidixic acid, which differs from the others for not containing a fluorine atom, had no effect on pP1192R activity at the concentrations tested. A summary of the results obtained with topoisomerase inhibitors is presented on Table 8.

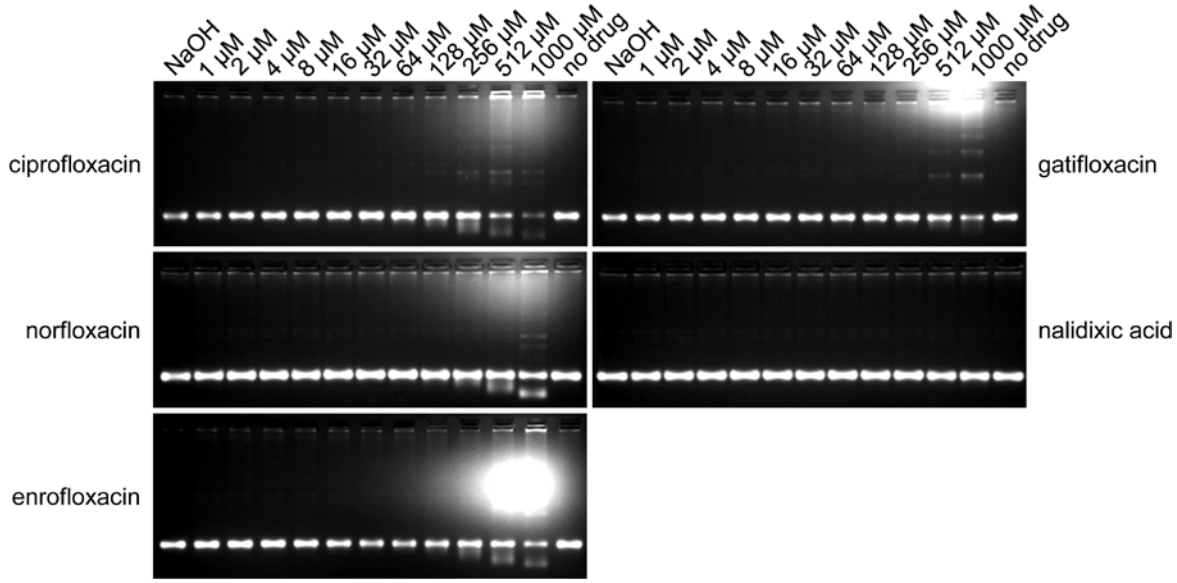


**Figure 56.** Inhibition of decatenation activity of pP1192R by quinolones.
*In vitro* decatenation reactions were performed essentially as described in section 4.2.21, using kDNA as substrate and optimal conditions defined for pP1192R activity. Reactions were incubated at 37 ºC for 30 minutes, after which they were stopped by addition of 4 µl of stop solution. The final drug concentration used in each reaction is indicated at the top of each lane. Images are representative of, at least, two independent experiments.

**Table 8.** Lowest effective doses for topoisomerase inhibitors tested in this work.

| Drug | Class | Category | Target | Lowest effective dose |
|---|---|---|---|---|
| coumermycin A1 | aminocoumarins | inhibitor | ATP-binding domain | ≥ 1 µM (1.11 µg/ml) |
| doxorubicin | anthracyclines | poison/inhibitor | topoII | ≥ 1 µM (0.58 µg/ml) |
| etoposide | podophyllotoxins | poison | topoII | ≥ 256 µM (150.67 µg/ml) |
| genistein | isoflavones | poison | not determined | ≥ 32 µM (8.65 µg/ml) |
| *m*-AMSA | aminoacridines | poison | topoII | ≥ 8 µM (3.44 µg/ml) |
| naringenin | flavanones | inhibitor | not determined | ≥ 128 µM (34.85 µg/ml) |
| nalidixic acid | quinolones | poison | topoII | not detected |
| ciprofloxacin | fluoroquinolones | poison | topoII | ≥ 128 µM (42.41 µg/ml) |
| norfloxacin | fluoroquinolones | poison | topoII | ≥ 256 µM (81.75 µg/ml) |
| enrofloxacin | fluoroquinolones | poison | topoII | ≥ 256 µM (92.00 µg/ml) |
| gatifloxacin | fluoroquinolones | poison | topoII | ≥ 512 µM (206.04 µg/ml) |
| topotecan | camptothecins | inhibitor | topoI | not detected |

## 4.4 Discussion

Heterologous expression of P1192R was attempted in three different expression systems: *E. coli*, *P. pastoris* and *S. cerevisiae*. Production of recombinant pP1192R was successful in the first and last systems, but we were unable to express this ASFV ORF in *Pichia pastoris* and the reasons behind this failure were never fully addressed. We confirmed integration of the DNA construct in the genome of X33 cells but, although we never detected expression of the recombinant protein, we did not analyze transcription of P1192R upon induction of expression. Therefore, we cannot conclude if the ORF was efficiently transcribed and the problem resided in the translation of the corresponding mRNA, or if simply no (complete) transcript was originated from P1192R. Still, premature termination of transcripts has been described for *P. pastoris* (Scorer, Buckholz, Clare, & Romanes, 1993; Withers-Martinez *et al.*, 1999), being considered a great handicap for this expression system, and usually occurs due to the presence of AT-rich regions in the sequences of the genes being expressed. ORF P1192R has a high AT content, of 52.75%, as indicated on Chapter 2, and it also contains stretches of AAs and TTs along its sequence, some resembling known yeast consensus sequences for 3' end formation in mRNAs (Henikoff & Cohen, 1984). If termination of transcription is predicted to occur at these sites, ending at nucleotides 1137 or 2291, translation of the respective mRNAs is predicted to originate proteins containing a maximum of 379 or 763 amino acid residues, with 42.15 or 85.56 kDa, respectively. These predicted peptide masses are similar to those of intense bands (48 and 75 kDa) visible in total extracts of induced *P. pastoris* cultures which were supposed to express P1192R (data not shown). Therefore, our hypothesis is that absence of expression of P1192R in this system was due to premature termination of its transcription. Curiously, this phenomenon is also described to occur in *S. cerevisiae* (Henikoff & Cohen, 1984; Romanos *et al.*, 1992) but we easily expressed P1192R in this yeast. This difference between the two expression systems is not uncommon, and a similar case was observed for the expression of the HIV-1 envelope protein (Scorer *et al.*, 1993), which was found to contain sequences causing premature polyadenylation in *P. pastoris* but not in *S. cerevisiae*.

In the *E. coli* system the problem did not reside in expression of P1192R, which was rapidly achieved and in considerable amounts, but the solubilization and purification steps were challenging. We could not achieve solubilization of the protein under native conditions since it appeared to accumulate in inclusion bodies. *E. coli* strains C41(DE3) and C43(DE3) have been shown to be resistant to the expression of toxic proteins (Miroux & Walker, 1996) and this resistance was determined to originate in a lower transcription rate of the exogenous genes, due to mutations in the promoter driving the expression of the T7 RNA polymerase

(Wagner *et al.*, 2008). Since rapid, high-level expression of proteins is thought to be one of the causes of accumulation of proteins in inclusion bodies (Rosano & Ceccarelli, 2014), use of strains that express recombinant proteins at lower rates for expression of pP1192R could have avoided its accumulation in inclusion bodies and improved its solubility. However, use of C41(DE3) and C43(DE3) was unsuccessful, as no expression of pP1192R was detected. Thus, we had to resort to the use of chaotropic agents or detergents for efficient extraction of the recombinant protein from the inclusion bodies. Two of the detergents tested, sarcosyl and fos-choline-12, were found to efficiently promote the solubilization of pP1192R. Curiously, both detergents have been used for the extraction and purification of transmembrane proteins (Leviatan, Sawada, Moriyama, & Nelson, 2010), and our bioinformatics analysis (reported in Chapter 2) predicted the existence of a transmembrane domain in pP1192R. Nevertheless, none of these detergents allowed for purification of functionally active protein, most likely because we could never fully remove them from solution. Addition of Triton X-100 up to 2% to the lysis buffer for extraction of recombinant pP1192R expressed in *E. coli* had no effect on the solubility of the protein. However, addition of only 0.1% of the same detergent to the yeast lysis buffer was sufficient to maintain pP1192R soluble without hindering its functionality, as demonstrated by activity assays. While in *E. coli* it is likely that the detergent is required to completely disaggregate the inclusion bodies, in which pP1192R is likely to have accumulated, in *S. cerevisiae* the detergent probably just stabilizes the protein, perhaps reducing protein-protein interactions, since it seems to counteract a tendency of the protein to aggregate. Therefore, it's logical for the strength of the detergent required in the first case to be higher than in the second case, explaining why Triton X-100 was not efficient until after solubilization of pP1192R. Interestingly, Triton X-100 is also widely used for the solubilization of membrane proteins (Kalipatnapu & Chattopadhyay, 2005).

A pairwise alignment of the L60 pP1192R and a C-terminally truncated (at residue 1186) yeast Top2p revealed that they are only 25.6% identical (335 identical amino acid residues), with a similarity of 42.2% (551 amino acid residues). With such low identity we were expecting to obtain only a partial rescue of the mutant phenotype in yeast complementation assays, as was the case (Figure 42). We also observed a slight reduction in the growth of yeast cells expressing pP1192R fused with the C-terminal domain of Top2p at the permissive temperature. This suggests that expression of this protein in high quantities is toxic for the cell, which is not uncommon when overexpressing topoisomerases in yeast (Bjornsti & Fertala, 1999; Lindsley, 1999), or that the presence of the viral protein in the cell nucleus may hinder the activity of Top2p, perhaps through the formation of less-functional heterodimers. Unfortunately, the *in vitro* activity assay using yeast soluble protein extracts does not provide

any insight into this matter because, even though pP1192R seems to be more active than pP1192R-Top2pCterm, their concentration in the extracts doesn't appear to be similar, as judged by Coomassie staining of SDS-PAGE gels, and that could explain the difference observed. We also cannot fully explain the difference between rescue in cells grown in plates and in liquid media. In liquid media the nutrients surround the cells and are, therefore, readily available for uptake, while in solid media their uptake is conditioned, especially after several cell divisions, and this may certainly influence cellular growth rate. Also, it takes 24 - 48 hours for wild-type cells to form colonies in solid media. Since the pP1192R-TOP2Cterm-expressing cells already present a reduction in growth, the combination of these factors may explain the difference observed between rescue in solid and in liquid media.

Our indications that the activity observed when we tested for pP1192R activity in the *in vitro* decatenation assays using yeast soluble protein extracts is of a type II DNA topoisomerase are two-fold: first, this class of enzymes requires ATP to function and activity is not observed when ATP is omitted from the reaction buffer; second, total extracts of JCW26 cells transformed with the empty pYES2 vector provide the background level of DNA and RNA contamination from the extract as well as residual decatenation activity from the endogenous temperature-sensitive Top2p enzyme or activities from other DNA-acting enzymes, and there is a clear difference between these lanes (Figure 44, upper panel, lanes 1 - 5) and those in which decatenation is observed. It is of relevance to consider that, according to the supplier of the kinetoplast DNA, the kDNA used in the assay generally contains some level of open circular (nicked) DNA, with the nicking not being caused by the tested topoII, but, despite that, a topoisomerase I is not expected to significantly decatenate the network of catenated minicircles. In these assays we also found that the highest volume used of extract of pP1192R-expressing cells would result in an accumulation of a considerable amount of kDNA in the well of the agarose gel (Figure 44, lane 10). Even though this could appear to be a result of lack of activity, to which the contents of the total extract could have somehow contributed, it is also possible that this is a result of excess of activity, as it has been described that high concentrations of topoII may catalyze the catenation of circular DNA molecules, mimicking lack of activity (Nitiss *et al.*, 2012). Indeed, the results obtained when using more than 50 nM of purified pP1192R (Figures 46 and 53) demonstrate that it's a topoII concentration-dependent effect.

Purifying soluble pP1192R allowed us to characterize its activity and the optimal conditions for it. We found that these do not differ significantly from the optimal conditions for other well-studied type II DNA topoisomerases (see Table 9), but there is an interesting

**Table 9.** Comparison of optimal conditions for purified eukaryotic type II topoisomerases.

| | **ASFV** | **PBCV-1** | ***S. cerevisiae*** | ***Drosophila*** | ***H. sapiens*** | ***Arabidopsis*** |
|---|---|---|---|---|---|---|
| pH | 7.5 | 8.5 | 7.4 | 7.9 | 7.5 / 8.0 | 8.0 |
| salt (mM) | 75 | 125 | 100 - 165 | 100 | 150 | 125 |
| $MgCl_2$ (mM) | 6 | 2.5 | 5 / 10 | 5 | 10 | 5 - 7.5 |
| ATP (mM) | 2.0 | 1 | 0.5 - 1 | 1 | 1 / 2 | 1 |
| DTT / β-SH (mM) | 1 | 0 | 1 / 5 | 0.01 | 0.5 / 1 | 0.5 |
| Temperature ($^oC$) | 30 - 37 | 25 | 30 - 35 | 30 | 37 | 20 - 25 |

difference. pP1192R has a very low salt requirement, with 75 mM being optimal and its activity is strongly diminished at salt concentrations which are optimal for nuclear topoisomerases. This may be due to the cellular context in which the enzymes are expected to work, and viral pP1192R is likely to function in a cytoplasmic compartment in which the salt concentration is lower than in the nucleus. However, the topoII from the *Paramecium bursaria Chlorella* virus 1 (PBCV-1), which is a NCLDV like ASFV, has higher salt requirements than pP1192R. We also found that the activity of pP1192R is inhibited by common topoII inhibitors, with coumermycin A1, doxorubicin and *m*-AMSA being the ones for which an effect is observed with very low doses. Even though use of coumermycin A1 in cancer therapy has been abandoned since it presents, like other aminocoumarins, high *in vivo* toxicity, low water solubility and low effectiveness (Pommier *et al.*, 2010), *m*-AMSA and doxorubicin are widely used in anti-cancer treatments. While there are no activity studies concerning the type II topoisomerases from the domestic pig, the amino acid sequences of the topoIIα from *Sus scrofa* and its homolog from *H. sapiens* share 94.3% of identity and 96.4% of similarity, with most of the differences being located to the C-terminal region, and therefore we expect topoII inhibitors to act identically on both enzymes. Still, inhibitory values obtained from *in vitro* studies are frequently divergent and are often obtained using plasmid relaxation assay. Therefore, to evaluate the possibility of using topoII inhibitors for the treatment of ASFV infection, studies on the *Sus scrofa* type II DNA topoisomerases should be performed as to establish proper comparisons.

## 4.5 Bibliography


Adachi, N., Miyaike, M., Ikeda, H., & Kikuchi, A. (1992). Characterization of cDNA encoding the mouse DNA topoisomerase II that can complement the budding yeast top2 mutation. *Nucleic Acids Research*, *20*(20), 5297–5303.

Anderson, V. E., & Osheroff, N. (2001). Type II topoisomerases as targets for quinolone antibacterials: turning Dr. Jekyll into Mr. Hyde. *Current Pharmaceutical Design*, *7*(5), 337–53.

Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M. J., & Walker, J. E. (2000). Characterisation of new intracellular membranes in Escherichia coli accompanying large scale over-production of the b subunit of F1Fo ATP synthase. *FEBS Letters*, *482*(3), 215–219.

Arnone, J. T., Walters, A. D., & Cohen-Fix, O. (2013). The dynamic nature of the nuclear envelope: Lessons from closed mitosis. *Nucleus*, *4*(4), 261–266.

Austin, C. A., Marsh, K. L., Wasserman, R. A., Willmore, E., Sayer, P. J., Wang, J. C., & Fisher, L. M. (1995). Expression, Domain Structure, and Enzymatic Properties of an Active Recombinant Human DNA Topoisomerase IIbeta. *Journal of Biological Chemistry*.

Bailly, C. (2001). DNA relaxation and cleavage assays to study topoisomerase I inhibitors. In M. J. W. B. T.-M. in E. Jonathan B. Chaires (Ed.), *Drug-Nucleic Acid Interactions* (Vol. Volume 340, pp. 610–623). Academic Press.

Baron, M., Reynes, J.-P. P., Stassi, D., & Tiraby, G. (1992). A selectable bifunctional β-galactosidase::phleomycin-resistance fusion protein as a potential marker for eukaryotic cells. *Gene*, *114*(2), 239–243.

Baviskar, A. T., Madaan, C., Preet, R., Mohapatra, P., Jain, V., Agarwal, A., Guchhait, S. K., Kundu, C. N., Banerjee, U. C., & Bharatam, P. V. (2011). N-fused imidazoles as novel anticancer agents that inhibit catalytic activity of topoisomerase IIα and induce apoptosis in G1/S phase. *Journal of Medicinal Chemistry*, *54*(14), 5013–5030.

Baylis, S. A., Dixon, L. K., Vydelingum, S., & Smith, G. L. (1992). African swine fever virus encodes a gene with extensive homology to type II DNA topoisomerases. *Journal of Molecular Biology*, *228*(3), 1003–1010.

Bhoo, S. H., Lai, H., Ma, J., Arntzen, C. J., Chen, Q., & Mason, H. S. (2011). Expression of an immunogenic Ebola immune complex in Nicotiana benthamiana. *Plant Biotechnology Journal*, *9*(7), 807–816.

Biersack, H., Jensen, S., & Westergaard, O. (1999). Rapid purification of DNA topoisomerase II containing a hexahistidine tag by metal ion affinity chromatography. *Methods in Molecular Biology (Clifton, N.J.)*, *94*, 235–41.

Bjornsti, M.-A., & Fertala, J. (1999). Overexpression and Purification of DNA Topoisomerase I from Yeast. In M.-A. Bjornsti & N. Osheroff (Eds.), *DNA Topoisomerase Protocols SE - 19* (Vol. 94, pp. 179–186). Humana Press.

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., & Boeke, J. D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)*, *14*(2), 115–32.

Bretthauer, R. K., & Castellino, F. J. (1999). Glycosylation of Pichia pastoris-derived proteins. *Biotechnology and Applied Biochemistry*, *30 ( Pt 3)*, 193–200.

Caron, P. R., Watt, P., & Wang, J. C. (1994). The C-terminal domain of Saccharomyces cerevisiae DNA topoisomerase II. *Molecular and Cellular Biology*, *14*(5), 3197–207.

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in Microbiology*, *5*, 63.

Couderc, R., & Baratti, J. (1980). Oxidation of Methanol by the Yeast, Pichia pastoris. Purification and Properties of the Alcohol Oxidase. *Agricultural and Biological Chemistry*, *44*(10), 2279–2289.

Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P., & Stillman, C. A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Molecular and Cellular Biology* , *9* (3 ), 1316–1323.

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 634–643.

Davies, J., & Jimenez, A. (1980). A New Selective Agent for Eukaryotic Cloning Vectors. *The American Journal of Tropical Medicine and Hygiene* , *29* (5 Part 2 ), 1089–1092.

Dickey, J. S., Choi, T.-J., Van Etten, J. L., & Osheroff, N. (2005). Chlorella virus Marburg topoisomerase II: high DNA cleavage activity as a characteristic of Chlorella virus type II enzymes. *Biochemistry*, *44*(10), 3899–908.

Donovan, R. S., Robinson, C. W., & Glick, B. R. (1996). Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter. *Journal of Industrial Microbiology*, *16*(3), 145–154.

Drocourt, D., Calmels, T., Reynes, J.-P. P., Baron, M., & Tiraby, G. (1990). Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Research*, *18*(13), 4009–4009.

Drugmand, J.-C., Schneider, Y.-J., & Agathos, S. N. (2012). Insect cells as factories for biomanufacturing. *Biotechnology Advances*, *30*(5), 1140–1157.

Esposito, D., & Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*, *17*(4), 353–358.

Fortune, J. M., & Osheroff, N. (2000). Topoisomerase II as a target for anticancer drugs: when enzymes stop being nice. *Progress in Nucleic Acid Research and Molecular Biology*, *64*, 221–253.

García-Beato, R., Freije, J. M., López-Otín, C., Blasco, R., Viñuela, E., & Salas, M. L. (1992). A gene homologous to topoisomerase II in African swine fever virus. *Virology*, *188*(2), 938–47.

Gatignol, A., Baron, M., & Tiraby, G. (1987). Phleomycin resistance encoded by the ble gene from transposon Tn 5 as a dominant selectable marker in Saccharomyces cerevisiae. *Molecular and General Genetics MGG*, *207*(2-3), 342–348.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, *274*(5287), 546, 563–567.

Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in Escherichia coli. *The Protein Journal*, *32*(6), 419–25.

Gottesman, S. (1996). Proteases and their targets in Escherichia coli. *Annual Review of Genetics*, *30*(1), 465–506.

Grodberg, J., & Dunn, J. J. (1988). OmpT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology* , *170* (3 ), 1245–1253.

Haldane, A., & Sullivan, D. (2001). DNA Topoisomerase II-Catalyzed DNA Decatenation. In N. Osheroff & M.-A. Bjornsti (Eds.), *DNA Topoisomerase Protocols SE - 2* (Vol. 95, pp. 13–23). Humana Press.

Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T. a, Li, H., Choi, B.-K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., & Gerngross, T. U. (2006). Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science (New York, N.Y.)*, *313*(5792), 1441–1443.

Harju, S., Fedosyuk, H., & Peterson, K. R. (2004). Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnology*, *4*, 8.

Henikoff, S., & Cohen, E. H. (1984). Sequences responsible for transcription termination on a gene segment in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *4*(8), 1515–1520.

Higgins, D., Busser, K., Comiskey, J., Whittier, P., Purcell, T., & Hoeffler, J. (1998). Small Vectors for Expression Based on Dominant Drug Resistance with Direct Multicopy Selection. In D. Higgins & J. Cregg (Eds.), *Pichia Protocols SE - 4* (Vol. 103, pp. 41–53). Humana Press.

Higgins, D. R., & Cregg, J. M. (1998). Introduction to Pichia pastoris. *Methods in Molecular Biology*, *103*, 1–15.

Houdebine, L.-M. (2009). Production of pharmaceutical proteins by transgenic animals. *Comparative Immunology, Microbiology and Infectious Diseases*, *32*(2), 107–121.

Huang, W. M., Wei, L. S., & Casjens, S. (1985). Relationship between bacteriophage T4 and T6 DNA topoisomerases. T6 39-protein subunit is equivalent to the combined T4 39- and 60-protein subunits. *Journal of Biological Chemistry , 260* (15 ), 8973–8977.

Ito, H., Fukuda, Y., Murata, K., & Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *Journal of Bacteriology*, *153*(1), 163–8.

Jensen, S., Andersen, a H., Kjeldsen, E., Biersack, H., Olsen, E. H., Andersen, T. B., Westergaard, O., & Jakobsen, B. K. (1996). Analysis of functional domain organization in DNA topoisomerase II from humans and Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *16*(7), 3866–77.

Kalipatnapu, S., & Chattopadhyay, A. (2005). Membrane Protein Solubilization: Recent Advances and Challenges in Solubilization of Serotonin1A Receptors. *IUBMB Life*, *57*(7), 505–512.

Kushnirov, V. V. (2000). Rapid and reliable protein extraction from yeast. *Yeast (Chichester, England)*, *16*(9), 857–60.

Lavrukhin, O. V, Fortune, J. M., Wood, T. G., Burbank, D. E., Van Etten, J. L., Osheroff, N., & Lloyd, R. S. (2000). Topoisomerase II from Chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase. *The Journal of Biological Chemistry*, *275*(10), 6915–21.

Leviatan, S., Sawada, K., Moriyama, Y., & Nelson, N. (2010). Combinatorial Method for Overexpression of Membrane Proteins in Escherichia coli. *Journal of Biological Chemistry , 285* (31 ), 23548–23556.

Li, H., & d'Anjou, M. (2009). Pharmacological significance of glycosylation in therapeutic proteins. *Current Opinion in Biotechnology*, *20*(6), 678–684.

Lindsley, J. E. (1999). Overexpression and Purification of Saccharomyces cerevisiae DNA Topoisomerase II from Yeast. In M.-A. Bjornsti & N. Osheroff (Eds.), *DNA Topoisomerase Protocols SE - 20* (Vol. 94, pp. 187–197). Humana Press.

Lundblad, V. (2001). Yeast cloning vectors and genes. *Current Protocols in Molecular Biology*, *Chapter 13*(1992), Unit13.4.

Makarevitch, I., & Somers, D. a. (2005). Purification and characterization of topoisomerase IIA from Arabidopsis thaliana. *Plant Science*, *168*(4), 1023–1033.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., & Porro, D. (2012). Recombinant Protein Production in Yeasts. In A. Lorence (Ed.), *Recombinant Gene Expression SE - 17* (Vol. 824, pp. 329–358). Humana Press.

Maya, D., Quintero, M. J., de la Cruz Muñoz-Centeno, M., & Chávez, S. (2008). Systems for applied gene control in Saccharomyces cerevisiae. *Biotechnology Letters*, *30*(6), 979–87.

McClendon, A. K., Rodriguez, A. C., & Osheroff, N. (2005). Human topoisomerase IIα rapidly relaxes positively supercoiled DNA: Implications for enzyme action ahead of replication forks. *Journal of Biological Chemistry*, *280*(47), 39337–39345.

Miroux, B., & Walker, J. E. (1996). Over-production of Proteins in Escherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, *260*(3), 289–298.

Moffatt, B. A., & Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, *49*(2), 221–227.

Murray, C. J., & Baliga, R. (2013). Cell-free translation of peptides and proteins:from high throughput screening to clinical production. *Current Opinion in Chemical Biology*, *17*(3), 420–426.

Nitiss, J. L. (2009). Targeting DNA topoisomerase II in cancer chemotherapy. *Nature Reviews. Cancer*, *9*(5), 338–350.

Nitiss, J. L., Soans, E., Rogojina, A., Seth, A., & Mishina, M. (2012). Topoisomerase assays. *Current Protocols in Pharmacology*, *Chapter 3*, Unit 3.3.

Osheroff, N., Shelton, E. R., & Brutlag, D. L. (1983). DNA topoisomerase II from Drosophila melanogaster. Relaxation of supercoiled DNA. *The Journal of Biological Chemistry*, *258*(15), 9536–9543.

Peng, H., & Marians, K. J. (1993). Escherichia coli topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. *Journal of Biological Chemistry*, *268*(32), 24481–24490.

Pommier, Y., Leo, E., Zhang, H., & Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chemistry & Biology*, *17*(5), 421–33.

Romanos, M. A., Scorer, C. a, & Clare, J. J. (1992). Foreign gene expression in yeast: a review. *Yeast*, *8*(6), 423–88.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges . *Frontiers in Microbiology* .

Scorer, C. a, Clare, J. J., McCombie, W. R., Romanos, M. A., & Sreekrishna, K. (1994). Rapid selection using G418 of high copy number transformants of Pichia pastoris for high-level foreign gene expression. *Bio/technology (Nature Publishing Company)*, *12*(2), 181–184.

Scorer, C. A., Buckholz, R. G., Clare, J. J., & Romanes, M. A. (1993). The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast Pichia pastoris. *Gene*, *136*(1–2), 111–119.

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli Physiology in Luria-Bertani Broth . *Journal of Bacteriology* , *189* (23 ), 8746–8749.

Silverstone, A. E., Arditti, R. R., & Magasanik, B. (1970). Catabolite-Insensitive Revertants of Lac Promoter Mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *66*(3), 773–779.

Simons, J. P., Wilmut, I., Clark, A. J., Archibald, A. L., Bishop, J. O., & Lathe, R. (1988). Gene Transfer into Sheep. *Nat Biotech*, *6*(2), 179–183.

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41.

St John, T. P., & Davis, R. W. (1981). The organization and transcription of the galactose gene cluster of Saccharomyces. *Journal of Molecular Biology*, *152*(2), 285–315.

Studier, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, *219*(1), 37–44.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130.

Su, X., Schmitz, G., Zhang, M., Mackie, R. I., & Cann, I. K. O. (2012). Chapter One - Heterologous Gene Expression in Filamentous Fungi. In G. M. G. and S. S. B. T.-A. in A. Microbiology (Ed.), *Advances in Applied Microbiology* (Vol. Volume 81, pp. 1–61). Academic Press.

Sugiki, T., Fujiwara, T., & Kojima, C. (2014). Latest approaches for efficient protein production in drug discovery. *Expert Opinion on Drug Discovery*, *9*(10), 1189–1204.

Thill, G. P., Davis, G. R., Stillman, C., Holtz, G., Brierley, R., Engel, M., Buckholtz, R., Kinney, J., Provow, S., Vedvick, T., & Seigel, R. S. (1990). Positive and negative effects of multi-copy integrated expression vectors on protein expression in Pichia pastoris. *Proceedings of the 6th International Symposium on Genetics of Microorganisms, vol. II*, 477–490.

Tschopp, J. F., Brust, P. F., Cregg, J. M., Stillman, C. a., & Gingeras, T. R. (1987). Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. *Nucleic Acids Research*, *15*(9), 3859–3876.

Vashisht Gopal, Y. N., & Kondapi, A. K. (2001). Topoisomerase II poisoning by indazole and imidazole complexes of ruthenium. *Journal of Biosciences*, *26*(2), 271–276.

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O., & de Gier, J.-W. (2008). Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences*, *105*(38), 14371–14376.

Wasserman, R. A., Austin, C. A., Fisher, L. M., & Wang, J. C. (1993). Use of yeast in the study of anticancer drugs targeting DNA topoisomerases: expression of a functional recombinant human DNA topoisomerase II alpha in yeast. *Cancer Research*, *53*(15), 3591–6.

Withers-Martinez, C., Carpenter, E. P., Hackett, F., Ely, B., Sajid, M., Grainger, M., & Blackman, M. J. (1999). PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Engineering , 12* (12 ), 1113–1120.

Worland, S. T., & Wang, J. C. (1989). Inducible overexpression, purification, and active site mapping of DNA topoisomerase II from the yeast Saccharomyces cerevisiae. *The Journal of Biological Chemistry*, *264*(8), 4412–6.

Yurimoto, H., Oku, M., & Sakai, Y. (2011). Yeast methylotrophy: Metabolism, gene regulation and peroxisome homeostasis. *International Journal of Microbiology*.

Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances*, *30*(5), 1158–1170.

# CHAPTER 5

# CONCLUDING REMARKS

This work aimed to characterize ASFV ORF P1192R, described in 1992 as having homology to type II DNA topoisomerases (Baylis, Dixon, Vydelingum, & Smith, 1992; García-Beato *et al.*, 1992a). Despite its identification and the suggestion that type II topoisomerase-inhibiting drugs have an effect on ASFV (Salas, Kuznar, & Viñuela, 1983; Mottola *et al.*, 2013), proof of pP1192R functionality remained inexistent. Heterologous expression of P1192R using either prokaryotic or eukaryotic systems was proposed and/or attempted by different research groups but, to our knowledge, the proteins obtained were never shown to possess type II topoisomerase activity.

Our projected characterization extended to several fronts: 1) expand the bioinformatics analysis of P1192R, by searching for functional and regulatory residues, motifs and domains, and perform a deep phylogenetic study in the context of ASFV and of topoisomerases; 2) determine if P1192R is expressed upon viral infection and, if so, describe its timing of expression and localization; 3) demonstrate that P1192R indeed codes for a functional type II DNA topoisomerase and determine the optimal conditions for its activity. As often happens, not everything went according to plan, but the essence of the plan was fulfilled.

Indeed, herein we confirmed that P1192R codes for a functional type II DNA topoisomerase, whose functional domains and structure are conserved even though its overall sequence is highly dissimilar from other known type II topoisomerases. We determined that pP1192R is expressed during viral infection, after the onset of viral DNA replication, accumulating in viral factories over time, and this was observed both in Vero cells using the Vero-adapted Ba71V and in swine macrophages infected with the field isolate L60. Finally, we were able to purify recombinant pP1192R and characterize its activity *in vitro*, and found that its optimal conditions are not significantly different from other well characterized topoII and that its activity is inhibited by common topoII inhibitors.

## 5.1 Characterization of pP1192R

The bioinformatics analysis performed on the deduced amino acid sequence of P1192R aimed the identification of functional and regulatory domains and motifs. One result from the analysis was the prediction of signals for nuclear localization and nuclear export of the protein. Even though ASFV was initially described as having a complete cytoplasmic cycle (McAuslan & Armentrout, 1974), not only was it observed that the virus was unable to replicate in enucleated cells (Ortin & Viñuela, 1977) but now there is also accumulated evidence on the importance of the nucleus for ASFV replication (Tabarés & Sánchez Botija, 1979; García-Beato, Salas, Viñuela, & Salas, 1992b; Rojo, García-Beato, Viñuela, Salas, & Salas, 1999; Ballester *et al.*, 2010; Simões, Martins, & Ferreira, 2013). In addition, several

ASFV proteins have been described to localize in the host cell nucleus (Goatley *et al.*, 1999; Eulálio *et al.*, 2007; Silk, Bowick, Abrams, & Dixon, 2007; Nunes-Correia *et al.*, 2008) and there have also been reports of an initial phase of ASFV DNA replication inside of the host cell nucleus (Tabarés & Sánchez Botija, 1979; García-Beato *et al.*, 1992b; Rojo *et al.*, 1999; Ballester *et al.*, 2010). Therefore, since pP1192R was predicted to code for a viral type II DNA topoisomerase which could have a role in DNA replication, it was plausible to consider that it could shuttle between the nucleus and the cytoplasm of the host cells, as these are the sites in which replication of viral DNA occurs.

However, as our experiments clearly showed, the putative NLS and NES were false positives, as the protein was never detected inside the nucleus, even after treatment with leptomycin B, an inhibitor of the CRM-1-dependent export pathway. Furthermore, the distribution of the protein seems to be conserved between mammalian and yeast cells, as the absence of complementation of the yeast temperature-sensitive mutant with the wild-type pP1192R version, as well as expression of GFP-pP1192R in both cell types, showed. In the case of yeast cells, pP1192R was only observed in the nucleus when it was fused with the C-terminal region of the yeast topoII, in which nuclear localization signals have been identified. Thus, it seems logical to conclude that whatever process for which pP1192R is required must occur in the cytoplasm.

Of the functional residues and motifs identified in the bioinformatics analysis, the putative NLS and NES, as well as the predicted catalytic tyrosine, were experimentally tested. While deletion of the first ones did not change the cellular distribution of the protein, mutation of the tyrosine at position 800 to a phenylalanine completely abrogated the activity of pP1192R, demonstrating its importance in the catalytic cycle of the protein. This tyrosine is expected to perform a nucleophilic attack on the DNA phosphodiester bond, resulting in the cleavage of the DNA. The tyrosine then remains covalently bound to the extremity of the DNA strand, avoiding its recognition as DNA damage. Although not experimentally demonstrated, mutation of the tyrosine at position 800 probably impairs the nucleophilic attack, hindering the cut of the DNA molecule. Also expected to be important for the catalytic activity of the topoisomerase is the quinolone-resistance determining region, composed by the motif EGDSA, as are other residues along the protein sequence. Consequently, further functional characterization could include the determination of which residues in the catalytic pocket of pP1192R are essential for topoisomerase activity, either by binding to DNA or to a divalent ion.

## 5.2 The role of pP1192R during viral infection

Despite the *in vitro* characterization of pP1192R's activity, the precise role of this protein during viral infection remains undetermined. Our results, in agreement with results on the transcription of this ORF (García-Beato *et al.*, 1992a), show that pP1192R is only detectable at an intermediate to late phase of infection, suggesting that it may have a role in transcription during the cytoplasmic phase of ASFV infection. Alternatively, since we observed that expression of pP1192R requires the onset of viral DNA replication and because the protein accumulates in viral factories over time, pP1192R may also be an integral part of the ASFV replication machinery, resolving complex topological structures that arise during progression of the replisome. Analysis of the preference of pP1192R for positively or negatively supercoiled DNA may help clarify if there is a bias for its participation in either transcription or replication phenomena. Positively supercoiled DNA is frequently formed during DNA replication due to separation of the DNA strands by helicases, while the steady state of DNA molecules is usually negatively supercoiled as it is energetically more favourable and facilitates the opening of the double helix. Some type II topoisomerases, like the bacterial topoisomerase IV and human topoIIα, have been shown to have the ability to distinguish between the handedness of DNA supercoils, which is in accordance to their importance and cellular role(s) (Crisona, Strick, Bensimon, Croquette, & Cozzarelli, 2000; Stone *et al.*, 2003; McClendon, Rodriguez, & Osheroff, 2005; McClendon, Dickey, & Osheroff, 2006). Therefore, a functional preference of pP1192R towards overwound (+) or underwound (–) supercoiled DNA may hint for its purpose in infected cells.

pP1192R may also participate in genome segregation, by facilitating the separation (decatenation) of newly-replicated DNA molecules. Cellular type II topoisomerases have been shown to participate in DNA replication, unwinding the supercoils generated by strand separation and pre-catenane formation after DNA synthesis, in chromosome separation and segregation, by facilitating the decatenation of intertwined chromosomes, and in transcription, also by relaxing supercoils (Schoeffler & Berger, 2008; Nitiss, 2009; Pommier, Leo, Zhang, & Marchand, 2010). Therefore, and since ASFV does not code for a type I topoisomerase, pP1192R is expected to be important for some, if not all, of these aspects of the ASFV viral cycle. Studies on the type II DNA topoisomerase from *S. cerevisiae*, Top2p, revealed that while in the absence of this protein (for example, by using thermo-sensitive mutants) cells complete DNA replication but die after entering mitosis (Holm, Goto, Wang, & Botstein, 1985), in the presence of a catalytically inactive form of Top2p cells are unable to complete DNA replication at sites where two replication forks meet (Baxter & Diffley, 2008). Therefore, analysis of ASFV infection in the presence of a catalytically inert pP1192R or in

the absence of this viral protein would probably be enlightening. If pP1192R is as important for viral DNA replication as we hypothesize, a virus containing an inactive form of the protein would likely be unable to completely replicate its DNA, thereby either stalling in late phases of the viral cycle or producing viral particles without any DNA content. Alternatively, a virus deleted for pP1192R would possibly possess the ability to finish DNA replication while being unable to separate the replicated genomes, thereby generating viral particles without any DNA or with aberrant genomic copies. A common possible outcome for both scenarios would be the production of viral particles incapable of generating a fruitful infection.

In yeast and mammalian cells, which possess type I topoisomerases, some of the cellular functions of topoII can be complemented by action of topoI with exception for the final steps of DNA separation, which exclusively require a topoII. However, since ASFV only possesses a topoII, it seems likely that DNA replication could halt early in the process due to accumulation of excessive supercoiling. In this case, substitution of pP1192R for a topoI, for example from a poxvirus or a mimivirus, could help to understand at which point of ASFV infection a type II topoisomerase is really required. The identification of viral and/or cellular partners of pP1192R during the viral cycle could certainly contribute to the determination of its precise role(s) during viral infection.

Salas and colleagues (1983) observed that the RNA synthesis activity detected from ASFV virions is reduced in a dose-dependent manner by coumermycin A1, an inhibitor of type II topoisomerases, suggesting that pP1192R is an integrant part of the viral particle. Its presence in the virion could also be a result of, or a requirement for, the participation of pP1192R in the packaging of the viral genome, as has been suggested for the Mimivirus topoII (Chelikani, Ranjan, Zade, Shukla, & Kondabagil, 2014). In addition, the existence of pP1192R in the virion could suggest that it may have a role in the early stages of infection. The virion-associated RNA polymerase has been shown to be responsible for the transcription of early-expressed ASFV genes (Salas, Rey-Campos, Almendral, Talavera, & Viñuela, 1986), and if pP1192R is an integrant part of the viral particle it may also participate in that process. Some of our results may contribute to the hypothesis that pP1192R is a virion protein: 1) a transmembrane domain was repeatedly predicted to exist in its amino acid sequence; 2) solubilization of recombinant pP1192R was only achieved when using detergents commonly used to solubilize membrane proteins; 3) preliminary results (not shown) on the localization of GFP-pP1192R expressed in a COS-1-based stable cell line indicated that the protein is predominantly perinuclear, perhaps coinciding with membranous organelles like the ER and/or the Golgi complex, and in contrast to GFP alone which is spread throughout the

entire cytoplasm; 4) *in vitro* decatenation activity was observed in purified viral particles in preliminary decatenation assays (not shown). Consequently, integrating these four results with the recent demonstration that the inner envelope of the ASFV particle is formed from ER-derived membranes (Suarez *et al.*, 2015) leads to the possibility that pP1192R is a transmembrane protein, localized to ER-membranes that will later become part of the viral inner envelope, thereby carrying pP1192R to the virion. Its presence there may be required during packaging of the viral genome, to unpackage it just after infection or for replicative or transcriptional processes at initial moments of infection.

### 5.3 P1192R as a possible target for antiviral therapy and vaccine design

Considering that neither vaccines nor treatments are available to control African swine fever, its recent dissemination to Eastern Europe emphasizes the threat the disease poses to global pig husbandry. Prevention in countries free of ASF is mainly based on stringent import policies that aim to ensure that neither living infected animals nor infected animal products are introduced into ASF-free regions, while control and eradication involve the rapid killing and disposal of all animal on infected premises, as well as restriction of animal movements and marketing ban among others. Ever since the disease was first described by Montgomery in 1921 (Montgomery, 1921), a considerable effort has been made to obtain a vaccine. The observation that survivors of ASFV infection can resist challenge by related virulent viruses (DeTray, 1957; Malmquist, 1963; Mebus & Dardiri, 1980) has nourished the prospect of obtaining an effective vaccine. However, the immunogenic formulations based on either inactivated virus (Stone & Hess, 1967; Mebus, 1987) or on the use of recombinant antigens (Neilan *et al.*, 2004) have shown to induce low or null protection levels. On the other hand, attenuated and/or naturally occurring low virulent ASFV isolates, although capable of conferring cross-protection against homologous isolates of high virulence (Stone, DeLay, & Sharman, 1968; Leitão *et al.*, 2001; Boinas, Hutchings, Dixon, & Wilkinson, 2004; Mulumba-Mfumu, Goatley, Saegerman, Takamatsu, & Dixon, 2015), have shown safety levels unacceptable for vaccine use. Recent reports on the deletion of ORF B119L (9GL) in the virulent isolates Malawi Lil/20/1 and Georgia2007/1 showed some promising results towards the construction of an effective vaccine (O'Donnell *et al.*, 2015a, 2015b), but the same had been shown previously (Lewis *et al.*, 2000) with no further developments. Alternatively, the knowledge gathered in recent years on pig immune responses against African swine fever virus (Takamatsu *et al.*, 2013), and on the viral resources to evade them (Correia, Ventura, & Parkhouse, 2013), allowed the design of approaches for the construction of mutant attenuated viruses deficient in their capacity to manipulate and evade the host

defense mechanisms. Nevertheless, up to now the large majority of the ASFV mutants obtained are not attenuated and those that have shown lower virulence, thus allowing for a challenge experiment, have shown low protection capacity and similar safety concerns when compared to attenuated and naturally occurring low virulent isolates (Abrams *et al.*, 2013). In this scenario, new strategies for the development of a vaccine are mandatory and a good option may reside on the creation of mutant viruses with self-limited replication capacity. ASFV is expected to be extremely independent of the host cell in terms of replicative and transcriptional machineries (Dixon, Chapman, Netherton, & Upton, 2013; Rodríguez & Salas, 2013), which not only implies that the ORFs coding for components of these machineries are essential but also increases the possibility for the manipulation of these aspects of the viral cycle, either through the use of drugs/inhibitors or by using genetic tools. Deletion of genes involved in these mechanisms may allow for the creation of an ASFV disabled infectious single cycle (DISC) virus. In theory, DISC viruses replicate only once in the natural target cells due to the lack of an essential gene product but can still trigger an immune response (Dudek & Knipe, 2006). Because of the deletion of the essential gene such viruses are incapable of producing infectious virions, unless the deleted function is provided by a complementing/helper cell line. Therefore, if pP1192R is as essential as we conjecture, deletion of ORF P1192R could be a step towards the production of a DISC vaccine.

On the other hand, our work demonstrated that the activity of pP1192R is inhibited by drugs such as doxorubicin, *m*-amsacrine or ciprofloxacin, which are commonly used in antitumoral or antibacterial treatments. In fact, type II topoisomerases are important targets for the control and elimination of cancer or of bacterial infections. Further *in silico* modeling of pP1192R's structure and docking studies using known structures for topoII inhibitors may help to uncover promising drugs that can eventually counteract the activity of the viral topoisomerase without hindering the cellular homologs. Therefore, and even though further experiments must be performed in order to evaluate the effects of topoisomerase inhibitors on ASFV infection, in cell culture and perhaps *in vivo*, if proven effective the use of topoII-inhibiting drugs could be a useful strategy for the control of ASF in case of an outbreak, avoiding the usually associated measures with large economic impact.

## 5.4 Bibliography


Abrams, C. C., Goatley, L., Fishbourne, E., Chapman, D., Cooke, L., Oura, C. A., Netherton, C. L., Takamatsu, H. H., & Dixon, L. K. (2013). Deletion of virulence associated genes from attenuated African swine fever virus isolate OUR T88/3 decreases its ability to protect against challenge with virulent virus. *Virology*, *443*, 99–105.

Ballester, M., Galindo-Cardiel, I., Gallardo, C., Argilaguet, J. M., Segalés, J., Rodríguez, J. M., & Rodríguez, F. (2010). Intranuclear detection of African swine fever virus DNA in several cell types from formalin-fixed and paraffin-embedded tissues using a new in situ hybridisation protocol. *Journal of Virological Methods*, *168*(1-2), 38–43.

Baxter, J., & Diffley, J. F. X. (2008). Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. *Molecular Cell*, *30*(6), 790–802.

Baylis, S. A., Dixon, L. K., Vydelingum, S., & Smith, G. L. (1992). African swine fever virus encodes a gene with extensive homology to type II DNA topoisomerases. *Journal of Molecular Biology*, *228*(3), 1003–1010.

Boinas, F. S., Hutchings, G. H., Dixon, L. K., & Wilkinson, P. J. (2004). Characterization of pathogenic and non-pathogenic African swine fever virus isolates from Ornithodoros erraticus inhabiting pig premises in Portugal. *The Journal of General Virology*, *85*(Pt 8), 2177–87.

Chelikani, V., Ranjan, T., Zade, A., Shukla, A., & Kondabagil, K. (2014). Genome Segregation and Packaging Machinery in Acanthamoeba polyphaga Mimivirus Is Reminiscent of Bacterial Apparatus. *Journal of Virology*, *88*(11), 6069–75.

Correia, S., Ventura, S., & Parkhouse, R. M. (2013). Identification and utility of innate immune system evasion mechanisms of ASFV. *Virus Research*, *173*(1), 87–100.

Crisona, N. J., Strick, T. R., Bensimon, D., Croquette, V., & Cozzarelli, N. R. (2000). Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements. *Genes & Development* , *14* (22 ), 2881–2892.

DeTray, D. E. (1957). Persistence of viremia and immunity in African swine fever. *American Journal of Veterinary Research*, *18*(69), 811–816.

Dixon, L. K., Chapman, D. A. G., Netherton, C. L., & Upton, C. (2013). African swine fever virus replication and genomics. *Virus Research*, *173*(1), 3–14.

Dudek, T., & Knipe, D. M. (2006). Replication-defective viruses as vaccines and vaccine vectors. *Virology*, *344*(1), 230–239.

Eulálio, A., Nunes-Correia, I., Salas, J., Salas, M. L., Simões, S., & Pedroso de Lima, M. C. (2007). African swine fever virus p37 structural protein is localized in nuclear foci containing the viral DNA at early post-infection times. *Virus Research*, *130*(1-2), 18–27.

García-Beato, R., Freije, J. M., López-Otín, C., Blasco, R., Viñuela, E., & Salas, M. L. (1992a). A gene homologous to topoisomerase II in African swine fever virus. *Virology*, *188*(2), 938–47.

García-Beato, R., Salas, M. L., Viñuela, E., & Salas, J. (1992b). Role of the host cell nucleus in the replication of African swine fever virus DNA. *Virology*, *188*(2), 637–649.

Goatley, L. C., Marron, M. B., Jacobs, S. C., Hammond, J. M., Miskin, J. E., Abrams, C. C., Smith, G. L., & Dixon, L. K. (1999). Nuclear and nucleolar localization of an African swine fever virus protein, I14L, that is similar to the herpes simplex virus-encoded virulence factor ICP34.5. *Journal of General Virology*, *80*(3), 525–535.

Holm, C., Goto, T., Wang, J., & Botstein, D. (1985). DNA topoisomerase II is required at the time of mitosis in yeast. *Cell*, *41*(2), 553–563.

Leitão, A., Cartaxeiro, C., Coelho, R., Cruz, B., Parkhouse, R. M., Portugal, F., Vigário, J. D., & Martins, C. L. (2001). The non-haemadsorbing African swine fever virus isolate ASFV/NH/P68 provides a model for defining the protective anti-virus immune response. *The Journal of General Virology*, *82*(Pt 3), 513–23.

Lewis, T., Zsak, L., Burrage, T. G., Lu, Z., Kutish, G. F., Neilan, J. G., & Rock, D. L. (2000). An African Swine Fever Virus ERV1-ALR Homologue, 9GL, Affects Virion Maturation and Viral Growth in Macrophages and Viral Virulence in Swine. *Journal of Virology*, *74*(3), 1275–1285.

Malmquist, W. A. (1963). Serologic and immunologic studies with African swine fever virus. *American Journal of Veterinary Research*, *24*, 450–459.

McAuslan, B., & Armentrout, R. (1974). The Biochemistry of Icosahedral Cytoplasmic Deoxyviruses. In W. Arber, R. Haas, W. Henle, P. H. Hofschneider, J. H. Humphrey, N. K. Jerne, P. Koldovský, H. Koprowski, O. Maaløe, R. Rott, H. G. Schweiger, M. Sela, L. Syruček, P. K. Vogt, & E. Wecker (Eds.), *Current Topics in Microbiology and Immunology / Ergebnisse der Mikrobiologie und Immunitätsforschung SE - 4* (Vol. 68, pp. 77–105). Springer Berlin Heidelberg.

McClendon, a K., Dickey, J. S., & Osheroff, N. (2006). Ability of viral topoisomerase II to discern the handedness of supercoiled DNA: bimodal recognition of DNA geometry by type II enzymes. *Biochemistry*, *45*(38), 11674–80.

McClendon, A. K., Rodriguez, A. C., & Osheroff, N. (2005). Human topoisomerase IIα rapidly relaxes positively supercoiled DNA: Implications for enzyme action ahead of replication forks. *Journal of Biological Chemistry*, *280*(47), 39337–39345.

Mebus, C. A. (1987). African swine fever. *Advances in Virus Research*, *35*, 251–269.

Mebus, C. A., & Dardiri, A. H. (1980). Western hemisphere isolates of African swine fever virus: asymptomatic carriers and resistance to challenge inoculation. *American Journal of Veterinary Research*, *41*(11), 1867–1869.

Montgomery, R. E. (1921). On A Form of Swine Fever Occurring in British East Africa (Kenya Colony). *Journal of Comparative Pathology and Therapeutics*, *34*(0), 159–191.

Mottola, C., Freitas, F. B., Simões, M., Martins, C., Leitão, A., & Ferreira, F. (2013). In vitro antiviral activity of fluoroquinolones against African swine fever virus. *Veterinary Microbiology*, *165*(1-2), 86–94.

Mulumba-Mfumu, L. K., Goatley, L. C., Saegerman, C., Takamatsu, H.-H., & Dixon, L. K. (2015). Immunization of African Indigenous Pigs with Attenuated Genotype I African Swine Fever Virus OURT88/3 Induces Protection Against Challenge with Virulent Strains of Genotype I. *Transboundary and Emerging Diseases*, n/a–n/a.

Neilan, J. G., Zsak, L., Lu, Z., Burrage, T. G., Kutish, G. F., & Rock, D. L. (2004). Neutralizing antibodies to African swine fever virus proteins p30, p54, and p72 are not sufficient for antibody-mediated protection. *Virology*, *319*(2), 337–342.

Nitiss, J. L. (2009). DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews. Cancer*, *9*(5), 327–37.

Nunes-Correia, I., Rodríguez, J. M., Eulálio, A., Carvalho, A. L., Citovsky, V., Simões, S., Faro, C., Salas, M. L., & Pedroso de Lima, M. C. (2008). African swine fever virus p10 protein exhibits nuclear import capacity and accumulates in the nucleus during viral infection. *Veterinary Microbiology*, *130*(1-2), 47–59.

O'Donnell, V., Holinka, L. G., Krug, P. W., Gladue, D. P., Carlson, J., Sanford, B., Alfano, M., Kramer, E., Lu, Z., Arzt, J., Reese, B., Carrillo, C., Risatti, G. R., & Borca, M. V. (2015a). African Swine Fever Virus Georgia 2007 with a Deletion of Virulence-Associated Gene 9GL (B119L), when Administered at Low Doses, Leads to Virus Attenuation in Swine and Induces an Effective Protection against Homologous Challenge. *Journal of Virology* , *89* (16 ), 8556–8566.

O'Donnell, V., Holinka, L. G., Krug, P. W., Gladue, D. P., Carlson, J., Sanford, B., Alfano, M., Kramer, E., Lu, Z., Arzt, J., Reese, B., Carrillo, C., Risatti, G. R., & Borca, M. V. (2015b). Deletion of African Swine Fever Virus Georgia 2007 virulence-associated gene *9GL* ( *B119L* ) leads to virus attenuation in swine at low doses while inducing an effective protection against homologous challenge. *Journal of Virology*, (June), JVI.00969–15.

Ortin, J., & Viñuela, E. (1977). Requirement of cell nucleus for African swine fever virus replication in Vero cells. *Journal of Virology*, *21*(3), 902–905.

Pommier, Y., Leo, E., Zhang, H., & Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chemistry & Biology*, *17*(5), 421–33.

Rodríguez, J. M., & Salas, M. L. (2013). African swine fever virus transcription. *Virus Research*.

Rojo, G., García-Beato, R., Viñuela, E., Salas, M. L., & Salas, J. (1999). Replication of African swine fever virus DNA in infected cells. *Virology*, *257*(2), 524–36.

Salas, M. L., Kuznar, J., & Viñuela, E. (1983). Effect of rifamycin derivatives and coumermycin A1 on in vitro RNA synthesis by African Swine Fever Virus. *Archives of Virology*, *77*(1), 77–80.

Salas, M. L., Rey-Campos, J., Almendral, J. M., Talavera, A., & Viñuela, E. (1986). Transcription and translation maps of african swine fever virus. *Virology*, *152*(1), 228–240.

Schoeffler, A. J., & Berger, J. M. (2008). DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Quarterly Reviews of Biophysics*, *41*(1), 41–101.

Silk, R. N., Bowick, G. C., Abrams, C. C., & Dixon, L. K. (2007). African swine fever virus A238L inhibitor of NF-kappaB and of calcineurin phosphatase is imported actively into the nucleus and exported by a CRM1-mediated pathway. *The Journal of General Virology*, *88*(Pt 2), 411–9.

Simões, M., Martins, C., & Ferreira, F. (2013). Host DNA damage response facilitates African swine fever virus infection. *Veterinary Microbiology*, *165*(1-2), 140–7.

Stone, M. D., Bryant, Z., Crisona, N. J., Smith, S. B., Vologodskii, A., Bustamante, C., & Cozzarelli, N. R. (2003). Chirality sensing by Escherichia coli topoisomerase IV and the mechanism of type II topoisomerases. *Proceedings of the National Academy of Sciences , 100* (15 ), 8654–8659.

Stone, S. S., DeLay, P. D., & Sharman, E. C. (1968). The Antibody Response in Pigs Inoculated with Attenuated African Swine Fever Virus. *Canadian Journal of Comparative Medicine*, *32*(3), 455–460.

Stone, S. S., & Hess, W. R. (1967). Antibody response to inactivated preparations of African swine fever virus in pigs. *American Journal of Veterinary Research*, *28*(123), 475.

Suarez, C., Andrés, G., Kolovou, A., Hoppe, S., Salas, M. L., Walther, P., & Locker, J. K. (2015). African swine fever virus assembles a single membrane derived from rupture of the endoplasmic reticulum. *Cellular Microbiology*, n/a–n/a.

Tabarés, E., & Sánchez Botija, C. (1979). Synthesis of DNA in cells infected with African swine fever virus. *Archives of Virology*, *61*(1-2), 49–59.

Takamatsu, H. H., Denyer, M. S., Lacasta, A., Stirling, C. M. a, Argilaguet, J. M., Netherton, C. L., Oura, C. a L., Martins, C., & Rodríguez, F. (2013). Cellular immunity in ASFV responses. *Virus Research*, *173*(1), 110–121.